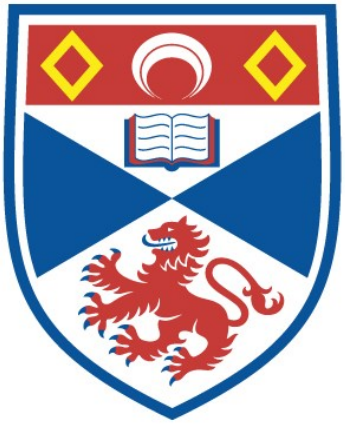
CAUSATION AND RESPONSIBILITY: FOUR ASPECTS OF THEIR RELATION

Horia Tarnovanu

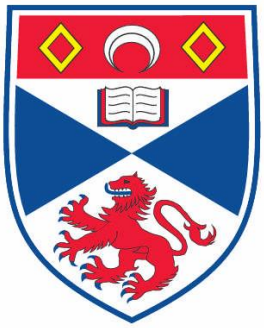
A Thesis Submitted for the Degree of PhD
at the
University of St Andrews



2015

Causation and Responsibility: Four Aspects of Their Relation

Horia Tarnovanu



This thesis is submitted in partial fulfilment for the degree of PhD at the University of St Andrews

Date of Submission: 25 September 2014

1. Candidate's declarations:

I, Horia Tarnovanu hereby certify that this thesis, which is approximately 70.000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student in September 2010 and as a candidate for the degree of PhD in September 2010; the higher study for which this is a record was carried out in the University of St Andrews between 2010 and 2014.

Date: 24 September 2014 Signature of candidate:

2. Supervisor's declaration:

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date: 24 September 2014 Signature of supervisor:

3. Permission for publication:

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

PRINTED COPY
a) No embargo on

ELECTRONIC COPY
a) Embargo on all or part of electronic copy for a period of 3 years on the following ground(s):
- Publication would preclude future publication

Supporting statement for electronic embargo request:
I would like to request an embargo on the electronic copy of the thesis for a period of 3 (three) years on the grounds that its online availability might preclude future publication. My intention is to turn the material included in the thesis into several papers in the next years, and early electronic access to the thesis in its current (unrevised) form might prevent me from achieving this aim.

Date: 24 September 2014 Signature of candidate: Signature of supervisor:

Abstract

The concept of causation is essential to ascribing moral and legal responsibility since the only way an agent can make a difference in the world is through her acts causing things to happen. Yet the extent and manner in which the complex features of causation bear on responsibility ascriptions remain unclear. I present an analysis of four aspects of causation which yields new insights into different properties of responsibility and offers increased plausibility to certain moral views.

Chapter I examines the realist assumption that causation is an objective and mind-independent relation between space-time located relata – a postulate meant to provide moral assessment with a naturalistic basis and make moral properties continuous with a scientific view of the world. I argue that such a realist stance is problematic, and by extension so are the views seeking to tie responsibility attributions to an objective relation.

Chapter II combines the context sensitivity of causal claims with the plausible idea that responsibility ascriptions rest on the assessment of causal sequences relating agents and consequences. I argue that taking context sensitivity seriously compels us to face a choice between moral contrastivism and a mild version of scepticism, *viz*. responsibility is not impossible, but ultimately difficult to identify with confidence. I show why the latter view is preferable.

Chapter III explores the concern that group agents would causally (and morally) overdetermine the effects already caused by their constituent individuals. I argue that non-reductive views of agency and responsibility lack a coherent causal story about how group agents impact the world as relatively independent entities. I explain the practical importance of higher-order entities and suggest a fictionalist stance towards group agency talk.

Chapter IV analyses the puzzle of effect selection – if causes have infinitely many effects, but only one or a few are mentioned in causal claims, what determines their selection from the complete set of consequents? I argue that the criteria governing the difference between effects and by-products lack clarity and stability. I use the concerns about appropriate effect selection to formulate an epistemic argument against consequentialism.

Acknowledgements

First of all, I would like to thank my supervisors, Rowan Cruft and Jessica Brown, for their patient guidance, stimulating comments, and many helpful reviews of my writing over these years.

I am also deeply indebted to Jonathan Schaffer and Ned Hall for extremely helpful conversations about and feedback on early drafts of some chapters in this dissertation.

I would also like to thank Stephen Yablo, Kent Hurtig, Peter Milne, Sarah Broadie, Sonia Roca-Royes, Simon Prosser, Ben Sachs, Adrian Haddock, Aaron Cotnoir, and Katherine Hawley for valuable feedback on my dissertation at various stages.

I am further grateful to audiences at the universities of Manchester, Kent, Salem, Illinois at Urbana-Champaign, Bordeaux, and Sassari for helpful feedback on my presentations. Thanks to Joseph Spino for commentaries on my paper in Illinois and to Gunnar Björnsson and Jon Williamson for our discussions at the conferences in Salem and Kent, respectively.

I would also like to thank the philosophy departments at Rutgers and Harvard for welcoming me as a visitor and for their amazing and inspiring philosophical atmosphere.

Very special thanks to my partner Laura for her love and support over the course of writing my dissertation, and to my family and friends for their encouragement and optimism.

I also kindly acknowledge the University of Stirling and the Horizon Programme for their financial support.

Contents

Introduction

This dissertation analyses several dimensions of the relation between causation and responsibility, a theme that figures prominently in legal and moral theory, with definite overtones in social and political philosophy. In a simple sense, causation matters to responsibility because the only way to make a difference in the world is through our causal powers. Virtually every area where agents have an impact is an area where responsibility claims are warranted. Of course, responsibility is usually also taken to depend on a host of other agency-related conditions such as intentionality, control, foreseeability, epistemic competence, psychological consistency, or a special moral sense. However, if attention is mainly restricted to the outcomes of an agent's acts, causation matters to responsibility in a significant way.

Historically, the relationship between causation and responsibility is emphasized with remarkable sophistication in several writings from the 5th century BC's Greece. An example is Antiphon's *Tetralogies*, a collection of forensic dialogues. The second tetralogy illustrates a famous case, also known from Plutarch's *Life of Pericles* (1984: 36): a young man practices javelin throwing in the Gymnasium, and at the moment he throws, a boy runs into the javelin's path, is hit, and instantly killed. Plutarch notes that Pericles and Protagoras debated for a long time whether the negligent boy, the javelin thrower, the javelin itself, or those organising the games should be held responsible for the death. The importance of causation emerges on both sides of the debate: the prosecution saw the young javelin-thrower as responsible, for had he not thrown the javelin the victim would not have been killed (3 3.10); he also caused disquiet not only to the dead boy, but also to the living, childless parents (3 1.2); although it was an accidental killing, 'the effects were the same as those of a wilful murder' (3 3.7); and even if both young men partially caused the accident, the boy has already paid a harsh price, while the javelin-thrower has not yet been punished (3 3.10). In reply, the defence also relies on causation: the young man was in fact 'prevented from hitting the target' (3 2.4) and 'made a hit, it is true, but killed no one' (3 2.3). In holding that the young man made a hit but killed no one, the defence does not conflate causal and moral responsibility, but makes the rather refined claim that one may cause a result without also causing what is logically entailed by it (or that an action can be dissociated from some of its consequences). For the things we do are sometimes the

subject of 'sheer necessity [which] can force all men to belie their nature in both word and deed' (3 3.1). A somewhat different example is noticed by Bernard Williams in Sophocles' tragedy *Oedipus at Colonus*. When Oedipus laments his deeds (the fortuitous fulfilment of the prophecy that he would kill his father Laius and marry his mother Jocasta) he strangely claims that he 'suffered those deeds more than acted them' (Sophocles 1984: 266-267), a remark which expresses the difficulty of assuming responsibility for what one has caused, irrespective of whether it was unintended, uncontrolled, or unforeseen. As Williams notes:

> The whole of the Oedipus Tyrannus, that dreadful machine, moves to the discovery of just one thing, that *he did it*. (…) [W]e understand the terror of that discovery (…) because we know that in the story of one's life there is an authority exercised by what one has done, and not merely by what one has intentionally done. (…) *What has happened to him, in fact, is that he has brought it about*. (Williams 1993: 69-70, emphasis in original).

Oedipus' words show that caused results may sometimes weigh more heavily in the assessment of responsibility than intention, control, or rational competence. This intimate connection between causation and responsibility in ancient Greek thought is also reinforced by their identical etymological roots. The Greek noun *aitia* (and its cognate adjectives *aitios* and *aition*) translate as both 'cause' and 'responsibility' (guilt, blameworthiness).[1] According to Hankinson (1998: 63), the initial connotation was 'responsibility', and the term gradually eroded and settled on the more technical sense of 'causation' we find for instance in Aristotle's works.

Fast-forwarding about 2400 years, the relation between causation and responsibility emerges in a variety of domains. The use of the notion of cause is essential to legal inquiry, as to establish legal responsibility in criminal law and torts one needs to show that the harm was brought about, induced, permitted or contributed to by the agent (Hart and Honoré 1985, Morse 2000, 2004, Moore 2009, Stapleton 2008, 2009, Schaffer 2010). The last decades have also seen an interest in the relation between causation and moral responsibility (Sartorio 2004, 2005a,b, 2006a,b, McGrath 2005, Cushman 2008, Driver 2008, Baumann 2011). Particular attention has been dedicated to the question whether responsibility entails and/or is grounded in causal sequences, to the causal status of omissions and their relation to responsibility, and, more generally, to the way in which various complex features of causation bear on the understanding of responsibility. Furthermore, causation matters to political theory, where the concepts

[1] Cf. the Online Liddell-Scott-Jones Greek-English lexicon at http://stephanus.tlg.uci.edu/lsj/#eid=2893&context=lsj

of agency and responsibility are extended to collective entities (French 1984, Pettit 1993, 2001, Miller 2007). Examining whether states, institutions or international organizations could be responsible agents represents an exercise which allows us to clarify complexities in policy-making in a world where calls to action come before the identification of adequate actors (Erskine 2003). If politics are permissively defined as 'the authoritative allocation of values for a society' (Easton 1953: 129), if such values are shaped and distributed by the participants in the political process, and if these participants are organizations such as governments, formal institutions, national corporations, interest groups etc., then causation and responsibility are essential for the political domain. In addition, causes are relevant to historians, as history is more than a list of chronological sequences, but a branch of knowledge which appeals to the concept of responsibility to evaluate the contributions made by historical figures to critical changes, victories or disasters, human flourishing or suffering (Fischer and Ravizza 1994). Causal analysis has also been reclaimed in International Relations (Kurki 2008), or discussed by applied ethicists working in business and assessing corporate action and social responsibility (Vogel 2006, May, Cheney, and Roper 2007, Crane, Matter, and Spence 2013).

But a remarkable omission across these debates is a profound, precise understanding of what causation actually is. My dissertation contributes to filling this gap. The explanation of the diversified nuances of responsibility emerging in the areas mentioned above ultimately hinges on the understanding of this highly complex and contested notion. Causation is problematic along several dimensions:

1. There are questions about the fundamentals of causal relations – e.g., what kind of objective reality causation has, how causal language connects to real-world phenomena, or whether causal facts are reducible to fundamental non-modal facts about our world and the laws of nature.
2. There are controversies about the basic principles of causal relations (Hall 2004):
Locality: causes connect to effects *via* spatio-temporal chains of causes.
Dependence: counterfactual dependence is sufficient for causation: *x* causes *y* if had x not occurred, *y* would not have occurred.
Transitivity: if *x* causes *y* and *y* causes *z*, then *x* causes *z* (a relation *R* is transitive iff $\forall(x, y, z)\ ((Rxy \& Rxz) \supset Rxz)$.
Omissions: causal relata may be negative events.

Intrinsicness: whether a sequence of events is causal depends on the intrinsic features, properties and relations of that sequence of events. Each of the principles listed has been subject to dispute.

3. There are questions about the nature of relata (events vs. facts), their individuation (coarse- vs. fine-grained), or their number (binary vs. quaternary). There are also questions about the nature of connection (process, probability), about the selection of causes vs. conditions, or about causal direction (temporal or otherwise) (Schaffer 2007).

4. There are questions about special cases – e.g., causal overdetermination (*x1* and *x2* causally overdetermine an effect *y* if both occur and each of them is in itself sufficient to cause *y*), or double prevention (for instance, if a boulder is dislodged and comes careering down the mountain towards a walker who ducks at the appropriate time, the duck prevents1 the collision, which, had it occurred, it would have prevented2 the walker's continued stride). Did the boulder cause the walker's survival? After all, it caused the duck, which caused the survival.

The technical challenges associated with most of the features of causal relations go deep. My dissertation is a contribution to the search for an adequate account of causation that would explain responsibility attributions. I present an analysis of several controversial aspects of causation which yields new insights into different properties of moral responsibility.

The focus is on moral responsibility because I believe that all the areas mentioned above (law, politics, history, etc.) are underpinned by a common, more fundamental, moral concern. The reason why causation is a determinant of responsibility in these domains is that it is a determinant of *moral* responsibility. One may be tempted to argue that moral responsibility is distinct and independent from legal liability, national responsibility, historical accountability, or corporate social answerability. But all these forms of responsibility are underpinned by the more fundamental idea of moral responsibility. To deny the fundamental status of moral responsibility is to argue that other forms of responsibility need not reflect moral considerations, which I find hardly plausible. At a minimum, the concern for responsibility may reflect and be replaced with a concern for explanation, that is, it may be argued that we are not really interested in the attribution of responsibility, but in a kind of austere account of the complex relations holding amongst events in different disciplines (e.g., history, law, or politics). However, it will become clear that a)

explanation and responsibility are closely related concepts, and b) history, law, or politics are hardly only about brute, sequenced causal relations amongst a range of neutral events.

My discussion will be centred on four independent but loosely connected contentious aspects of causation, relevant to particular moral responsibility properties or views:

1. The objective reality of causal relations, relevant to views of responsibility aiming to provide moral assessment with a naturalistic underpinning.
2. The context sensitivity of causal claims, relevant to the stability and success of moral responsibility attributions.
3. The notion of higher-level causation (an instance of causal overdetermination), relevant to the notion of collective moral responsibility.
4. The mechanisms of selection on the effect side of causal relations, relevant to evaluating the responsibility for consequences.

The choice of these four aspects is motivated by two reasons:

a) The connection between the chosen aspects of causation and moral responsibility is *underresearched*. The current philosophical literature is dominated by discussions about responsibility for omissions (Thomson 2003, Sartorio 2009a, 2010, Bernstein forthcoming), the distinction between doing and allowing (Moore 2009, Sartorio 2011), the problem of resultant luck[2] (Sartorio 2012), or the question whether moral responsibility entails or is grounded by causal responsibility (Sartorio 2004, 2007, Driver 2008). A somewhat different key aspect explored in the literature is the determination of free will by causal factors beyond our reach (Sartorio forthcoming). The relevance of causation to moral assessment also receives extensive attention in legal theory, under the plausible assumption that moral responsibility underpins legal responsibility (Hart and Honoré 1985, Moore 2009, Stapleton 2008, 2009). Legal theory contains a detailed study of many important problems of causation, including a meticulous analysis of causal language, discussions of transitivity, omissions or

[2] Resultant luck refers to the moral luck in the way our acts or projects turn out (Nagel 1979). Resultant luck is moral luck about results or consequences. For instance, if a drunk driver hits a child and a similarly drunk driver collides with a tree, the former is morally unlucky and the latter is morally lucky.

overdetermination, accounts of locality and dependence as grounds of responsibility, and other critical issues.

In contrast, the objective reality of causation is discussed in relation to the ontological image revealed by contemporary physics (Price and Corry 2007), or associated with the concept of a *de re* necessity binding causal relata (Chakravartty 2005, Beebee 2006), but not considered in relation to ethical theory. The context sensitivity of causal discourse is examined to establish if it receives a pragmatic or semantic interpretation (Schaffer 2012), but not in relation to moral responsibility attributions. Causal overdetermination receives a great deal of attention in the literature, but the same cannot be said about *constitutive* causal overdetermination as a problem for collective responsibility claims.[3] In fact, there is no refined, coherent causal story of how collectives impact the world as relatively independent entities. Finally, the mechanism of selection of the effect side is mentioned *en passant* as a counterpart of causal selection (Schaffer 2005: 352), but has not been thoroughly researched.

b) The chosen aspects are *consequential* in that they challenge assumptions or features of widely accepted views of responsibility. Important questions arise about the naturalist underpinning of moral responsibility assessment; about the possibility of ultimate moral responsibility;[4] about the causal efficacy and blameworthiness of collective agents (in most social sciences); about the individuation of effects and the adequate scope of responsibility for consequences. Therefore, an accurate analysis of the selected aspects of causation will clarify some important features, views and problems of moral responsibility.

A note about methodology. The main methodological strategy I shall follow will be comparing intuitive verdicts about causation with those offered by different extant theories. In this sense, I subscribe to David Lewis' view that '[i]f an analysis of causation does not deliver the common-sense answer, that is bad trouble' (Lewis 1986b: 194). However, this does not mean that intuitions generate non-negotiable data; rather,

[3] Constitutive causal overdetermination occurs when both high-level or macro-causes (e.g., group agents) and low-level or micro-causes (e.g., the individual members group agents are constituted of) are in themselves sufficient to cause an effect.

[4] An agent A is ultimately responsible for an outcome O if O is entirely up to A or if A is *the* source or *the* cause of O, despite other factors (conditions) being involved in bringing O about. For instance, A is ultimately responsible for shooting B despite not being responsible for the manufacturing of the gun or for her own psychological constitution. Similar locutions are being truly, essentially, or fundamentally responsible. The concept of ultimate moral responsibility appears in Strawson (1994).

I take intuitions to be important guides to what we know about causation. I will carry out a conceptual analysis of several constraints on the concept of causation. I will not carry out a reductionist project aimed at showing how causal relations reduce to more basic ontological facts, i.e. fundamental physical states and fundamental physical laws. I will pay special attention to the places where our intuitions conflict.

As far as the concept of *responsibility* is concerned, several clarifying remarks are in order. It is a strenuous task to investigate the variety of contexts in which this elusive term appears. Following Velasquez (1985), responsibility could refer to:

a) a quality of the agent (e.g., an agent A is responsible in the sense that she is trustworthy or reliable),

b) a duty or obligation (e.g., An agent A is responsible for Public Relations), or

c) the attribution of an action and its outcomes (e.g., the captain was responsible for the shipwreck). This third sense (attribution to an agent A of an action φ and its outcomes ω) could be further analysed as

c1) Agent A caused a specific action φ,

c2) Agent A is responsible for paying damages arising from action φ, and

c3) Agent A intentionally φ-ed or brought ω about.

This dissertation will be mainly concerned with responsibility in the third sense (c), the attribution of an action and its outcomes to an agent.

Next, responsibility may attach to agents, actions or outcomes. The focus of the dissertation is on the responsibility for *outcomes* (events and states of affairs in the world). The responsibility for various outcomes of interest is the most popular and practical form of responsibility out there. Naturally, outcomes refer to an agent's acts and an agent's acts require a discussion of a host of other agency-related notions such as freedom, intention, or control; however, the analysis will be centred on causation. To do justice to all conditions and at the same time maintain the emphasis on causation, I will assume a composite conception of responsibility, *viz*. one that takes causation to be one of the core elements of responsibility, next to intention, control, adequate reasons, psychological consistency, or normative pressure towards an answer (for instance, the answer demanded by the legal norms). Since all these elements are important, responsibility is a composite notion.[5] However, the emphasis placed on its elements always varies: there is no such thing as an appropriate way of adjusting them,

[5] A similar view is suggested by Williams (1993).

one 'correct' conception of responsibility, but different interpretations based on how we balance its core elements. When outcome responsibility is under assessment, causation becomes the axis of evaluation rather than being demoted to the status of a mere secondary element, one amongst the many necessary conditions of responsibility. For further focus, I will also leave aside the views attaching responsibility to *beliefs* rather than choices, actions or results (e.g., Moya 2006).

As a further refinement, I will mainly discuss *moral* responsibility. Moral responsibility is usually distinguished from *causal responsibility,*[6] *strict* or *vicarious responsibility,*[7] *compensatory responsibility,*[8] and *legal responsibility*.[9] These terms often overlap. The dissertation is built around the idea that (outcome) moral responsibility *entails* causal responsibility and causal responsibility *grounds* moral responsibility (Sartorio 2007): since the only way to impact the world is through our causal powers, the moral assessment of our acts is intuitively related to what we cause.

[6] It is possible to ascribe causal responsibility without moral responsibility (when we hold the cat responsible for breaking the vase or the weather responsible for ruining our mountaineering trip), and moral responsibility without causal responsibility (for instance, suppose A cannot perform action φ unless B is babysitting for A. B's failure to help on this occasion does not directly cause A's failure to perform action φ, as A could have asked someone else to help; however, B could still feel morally responsible for A's failure to perform action φ).

[7] *Strict* or *vicarious responsibility* must also be ascribed without direct causal responsibility (for example, a parent may pay compensation for the actions of a child, or, analogously, a company may be responsible in the compensatory sense for the actions of an employee). This dissertation will mainly be concerned with moral responsibility for outcomes, where the causal component is essential.

[8] Moral responsibility is different from *compensatory responsibility* (paying damages arising from an action). In many cases, A is responsible in the compensatory sense if there is a legal doctrine L that A must pay damages to B and B is entitled to compensation from A in the specific situation S. If not legally specified, compensatory responsibility becomes more slippery a notion, given the gap between those responsible for paying damages arising from action A and those directly responsible for action A. However, a detailed analysis of this notion is beyond the scope of this dissertation. This idea brings out a further distinction, between moral responsibility and *legal responsibility* (cf. fn. 9).

[9] Legal responsibility is established in courts, related to formal sanctions and penalties, based on evidence and centred on the outcomes of an action. We say that X is legally responsible for an action A when she is susceptible to be penalized in the court system. In contrast, moral responsibility refers to the human conduct, prioritizes intentions over outcomes and focuses on our (informal) evaluative responses. Moral responsibility does not entail legal responsibility. Legal and moral responsibility frequently overlap (and we expect them to overlap in societies where laws uphold moral principles), but depart on various occasions. For instance, X is both morally and legally responsible for murdering Y, however, the mere intention to murder Y does not make X legally responsible, but undoubtedly morally questionable. Terms such as 'liability', 'accountability', or 'answerability' are frequently used as more legally-minded notions to refer to compensatory or legal responsibility (see for instance Duff 2007, 2009 for the distinction between responsibility as attributability referring to the connection between agents and outcomes, and responsibility as answerability referring to the connection between agents and their moral community). However, many authors use these terms interchangeably; some distinguish 'liability' from 'legal responsibility' (Cane 2002), some use 'accountable' to refer to compensatory responsibility (Pettit 2007), and others use 'accountability' and 'liability' to refer to different senses of moral responsibility (Watson 1996, Corlett 2001, respectively).

Although my principal emphasis will be on the relation between causal and moral responsibility, I will sometimes discuss the notion of legal responsibility.

Lastly, I will not touch on issues such as the concept of agent causation or the relation between causal determinism and moral responsibility – that is, I will leave aside the causation-related questions debated in the literature on action and free will (e.g., Mele 2009).

Here is a brief summary of my arguments:

Chapter I

In 'Causation in Fact, Causation to Us', I examine the realist assumption that causation is an objective and mind-independent relation between space-time located relata – a postulate meant to provide moral assessment with a naturalistic basis and make moral properties continuous with a scientific view of the world. I argue that such a realist stance is problematic, and by extension so are the views seeking to tie responsibility attributions to an objective relation. My strategy is to show that causation is based on incompatible intuitions and the best attempts to explain them fail to secure a robust sense of realism. Two parallel attempts are diagnosed: one defends a sense of objectivity through a reconciliation of our incompatible intuitions; the other conserves a different sense of objectivity within the structural equations framework. However, it emerges that both describing and modelling of causal facts are affected by deep-seated ambiguities, and as a result, realists needs to retreat towards more modest commitments. *Per extensionem*, the theories tying responsibility attributions to the evaluation of genuine causal sequences turn out to be fundamentally misguided. This happens mainly because establishing whether there are genuine causal sequences in the relevant sense turns out to be a difficult task.

Chapter II

In 'Moral Lessons from the Context Sensitivity of Causal Claims', I examine how the context sensitivity of causal claims impacts moral assessment in complex situations and argue that taking context sensitivity seriously generates important worries about ultimate moral responsibility. I take a look at some alternative devices for stabilising causal thinking in moral contexts and show why their applicability is in principle restricted. Lastly, I contend that we end up with a choice between moral contrastivism and a mild version of scepticism about moral responsibility, *viz*. moral responsibility is

not impossible, but ultimately difficult to identify with confidence. I show why the latter view is more plausible.

Chapter III

In 'Higher-Level Causation', I explore the concern that group agents would causally overdetermine the effects already caused by their constituent individuals (and thus would morally overdetermine them). I show that non-reductive views about collective entities need a coherent causal story independently of whether the overdetermination objection is decisive or not. In particular, I argue that non-reductive manoeuvres generate a degree of commitment to an independent source of causal efficacy and any attempt to reroute it through individuals in virtue of constitution would entail a breakdown of higher-level performance and effectiveness. In different terms, because one of the important conditions of agency (next to having representational and motivational states) is to be causally efficacious, any non-reductive view would have to show how groups impact the world as relatively independent agents, irreducible to the individuals constituting them. I draw conclusions about the appropriate level of responsibility ascriptions and the right stance one should take towards group agency talk.

Chapter IV

In 'Effect Selection', I examine the mechanism of selection on the effect side: if causes have infinitely many effects, but only one or a few are selectively highlighted and mentioned in causal claims, what determines their selection from the complete set of consequents? I scrutinise the difference between effects and by-products, side-effects, long term effects, end-results etc. and argue for the need to find a way between two unappealing extremes: (a) the claim that effects and by-products are metaphysically distinct; (b) the claim that there is no sense in which effects and by-products are objectively different and selection is always governed by context-dependent pragmatics. I point out a few uncharted asymmetries between causal selection and effect selection, and examine whether the two puzzles can receive a parallel treatment. Finally, I consider how effect selection may be relevant to determining the appropriate scope of responsibility for consequences, and develop a new epistemic argument against consequentialism.

I. Causation in Fact, Causation to Us

'[C]ausation (…) is not merely as Hume says, *to us*, but also *in fact*, the cement of the universe.' (Mackie 1980: 2)

Introduction

A number of philosophers have recently argued that moral responsibility for how things turn out in the world rests in important respects on causal sequences (Sartorio 2007, Driver 2008, Moore 2009). Since causation is the only way an agent can make a difference in her external environment, causation grounds moral responsibility in the sense that it determines, explains, or makes true responsibility claims.[10] For instance, the crew of the MV Sierra, a vessel hunting throughout the Atlantic, are responsible for pirate whaling because they actually harpooned critically endangered species without license and regardless of season. Most theorists interested in how causation and responsibility relate have tacit or explicit realist sympathies, i.e. they take causation to be an objective and mind-independent relation between space-time located relata. A key advantage of a realist stance about causation is that it provides moral assessment with a naturalistic basis, making moral properties continuous with a scientific view of the world: it is in virtue of a determinate, objective feature of reality that we take responsibility claims to be authentic. Thus Moore notes:

> If moral responsibility does not depend (in part) on causal responsibility, (…) then on what natural property does the moral property (of responsibility) depend? Surely on some; moral properties unrelated to natural properties would make for a very odd ontology, a 'non-naturalism' in ethics that many, including myself find unacceptably counter to a scientific view of the world. (…) It is better to think that 'cause' is univocal; it means the same thing in contexts attributing responsibility as in contexts of explanation: it refers to a natural relation that holds between events or states of affairs. Because moral responsibility is tied to such a natural relation, and because the law is tied to morality, the law also is tied to this natural relation. (Moore 2009: 4-5).

In this chapter, I argue that such a realist stance about causation is problematic, and by extension so are the views which hope to tie attributions of outcome responsibility to

[10] The grounding claim – captured by the 'responsibility rests on causation' slogan – needs to be distinguished from what is known in the literature as *the entailment claim* (Sartorio 2007, Driver 2008) – the idea that being responsible for an event or state of affairs entails causing it (exceptions notwithstanding). Note that the converse of the entailment claim is not true (causation does not entail responsibility).

an objective relation. I will not go as far as to claim that realism is incoherent and as a result one needs to reject most attributions of responsibility as groundless. But I aim to show that we have incompatible intuitions about the nature of causation and the best attempts to explain them fail to safeguard a robust form of realism. If my arguments are correct, then the theories making responsibility attributions dependent upon the evaluation of genuinely objective causal sequences will turn out to be fundamentally misguided. Prior to delving into details, two caveats need to be added. First, the focus of the chapter will be on causation and on the strategies defending its objectivity. It is only along the way that I mention how different pressures on the objectivity of causation affect moral evaluation. Second, the chapter relies on the intuitive idea that responsibility for outcomes rests on causation because the only way to make a difference in the world is through our causal powers. A full defence of the grounding/entailment relation between causation and responsibility will be presented in Chapter II.

Overview: In Section 1, I spell out two incompatible intuitions which I assume to be central to our conception of causation: (i) causation is an objective and mind-independent relation between events located in space-time; (ii) causation is tied to considerations of a perspectival, projected, or normative nature. I examine two ways of dealing with this incompatibility. One is to safeguard realism by taking causation to be objective and perspective-dependence to reflect the shiftiness of our causal talk (Section 2). The other is to endorse a different sense of objectivity by taking the structural equations framework to deliver an adequate and non-arbitrary encoding of causal relations amongst phenomena (Section 3). I argue that both describing and modelling of causal facts are affected by deep-seated ambiguities, and as a result, realists need to retreat towards more modest or measured commitments. *En route*, I show how the various considerations affecting causal realism bear implications for moral assessment. I conclude by setting out the implications of my analysis (Section 4).

1. Realism and Perspective-Dependence

1.1 Incompatible Intuitions

Causation is characterised by conflicting intuitions:[11]

(i) *The realist intuition*: causation is a natural relation – out in the world, independent of our experience or perspective, connecting concrete events, and known a posteriori in the same way we come to know relations of temporal succession or spatial proximity (Menzies 2009). For instance, it is a completely objective matter that Aurora Borealis is caused by solar winds colliding with atoms in the thermosphere.

(ii) *The perspective-dependence intuition*: our intuitive conception of causation is tied to considerations of a perspectival, projected, or normative nature. Any time we mention direct, contributing, total, proportional, root, negative, triggering, predisposing, or proximate causes, we appeal to set contextual parameters or evaluative standards. As a result, most intelligible causation appears as causation *to us* rather than causation *in fact*. For instance, direct or contributing causes are *selected* by us based on our immediate aims and expectations – we say that the short circuit caused the fire because short circuits are unexpected events, but in fact a complex set of conditions are required for the fire to occur (dry timber, oxygen in the air, etc.)

Following the first intuition, I will be using the term 'causal realism' for the view that causal relations are objective and mind-independent.[12] Despite criticism,[13] realism remains one of the most prominent intuitions about the nature of causation.

[11] Accounts of causation are rife with intuitions competing for priority. Tensions arise with respect to its distinctive principles (intrinsicness and regularity pull in opposite directions, locality and dependence are irreconcilable), its special cases (e.g., negative or redundant causation), and its analysis (e.g., some instances of probabilistic relations appear as deterministic, intuitions diverge on whether theories should focus on type or token events, etc.). But this evasive terrain conceals an even deeper tension: it is not perfectly transparent if causation is a feature of a mind-independent world or contains – at least in part – something about us.

[12] As Mackie notes, 'It is one thing to ask what causation is "in the objects", as a feature of a world that is wholly objective and independent of our thoughts, another to ask what concept (or concepts) of causation we have, and yet another to ask what causation is in the objects so far as we know it and how we know what we do about it.' (Mackie 1980: 2).

[13] Causal realism was challenged by Hume (1739), disputed by some of his famous followers (Russell 1913, Wittgenstein 1922, Putnam 1984), and replaced by attenuated positions such as constructive empiricism – the view that science is limited to the truth about observable aspects of the world (van Fraassen 1980), projectivism – the view that we talk about the features of our thinking as if they were features of the world (Blackburn 1993, Spohn 1993), or structural realism – the view that science can only show what structures relate the entities in the world but nothing about those entities themselves (Esfeld 2009).

Causal realism may be taken to refer specifically to the idea that causal relations involve some sort of *de re* necessity binding relata together – it was causal necessity that was traditionally targeted by critics of realism and described as a mere projected figment of our psychology (Hume) or relocated from nature to a category of understanding (Kant). However, causal realism is not necessarily a view about necessity; it can stand on its own, as a position regarding the existence of an objective phenomenon in a mind-independent world. It is one thing to inquire whether causal efficacy lies in the objects or 'in the determination of the mind' (Hume 1739: 1.13.14). It is a somewhat different thing to ask whether causation is an extensional and perspective-independent feature of reality. Throughout the chapter, I will refer to the latter sense (for recent discussions of arguments about necessity, see e.g. Chakravartty [2005] or Beebee [2006]). I will also leave aside other interpretations of the term.[14] Lastly, I will be using the 'causation in fact', factual analysis' and 'causal facts' to refer to genuine extensional causal sequences (*as per* Mackie [1980]), and not to causation relating abstract and non-spatiotemporal facts (propositions), *as per* Mellor (1995, 2004).

Following the second intuition, causation is something at least in part sensitive to the way we understand the world. Certain care is needed not to state the trivial: in a basic sense, everything depends on our understanding, as any ontological claim lacking epistemological support would be a shot in the dark. What is meant is that certain *kinds* of causation, some of its *features*, and the *identity conditions* of most causes appear to be defined in connection to set contextual parameters or evaluative standards. Call these parameters or standards 'normative considerations', where the term 'normative' has the non-technical, generic meaning of being related to, establishing or deriving from a standard, model, or 'the way things should be'. Norms cover a large spectrum[15] and may influence what causes what independently or in concert. Dependence on a certain

[14] 'Causal realism' is a somewhat idiosyncratic notion. First, most realists are reductionists, but realism does not axiomatically entail an ontological reduction of causation to non-causal facts about what happens plus facts about the laws governing the physical world. For instance, if Strawson (1989: 84) understands causal realism as a reductive view – 'there is something about the fundamental nature of the world in virtue of which the world is regular in its behaviour', Tooley (1987: 246) understands realism as a non-reductive, primitivist view – 'the truth values of causal statements are not, in general, logically determined by non-causal [more fundamental] facts'. Second, 'causal realism' is also associated with the view that causation is a matter of powers/dispositional properties of things to affect other things (e.g., Mumford and Anjum 2011): if powers are real, then causation is a matter of how dispositional properties manifest themselves.

[15] Examples include statistical norms, moral and legal norms, policies or rules of etiquette, ideologies, norms of proper functioning, norms of proportionality, reasonable expectations, or mental models.

type or number of norms defines a particular perspective. I will next offer some examples of perspective-dependence.

1.2 Evidence for Perspective-Dependence

Perspective-dependence is supported by a range of examples:

a) Causal selection. We draw a distinction between causes and background conditions. If a fire occurs in a forest, we promote the lightning strike to the status of cause and demote the presence of oxygen to the status of a background condition. But if the fire occurs in the space shuttle Discovery's rear engine compartment, where precautions against oxygen leaks are taken, the presence of oxygen is promoted to the status of cause and other factors demoted to the status of background conditions. Selection is perspective-dependent because it is made against what we take to be the expected or reasonable course of events. The answer to 'What caused an outcome and why?' is relativized to and varies with what is implicitly assumed as invariant or reasonable when formulating the causal question. (Compare this with how perception works: attention tracks changes or deviations from specific invariant features of the world – what it takes to be stable, reliable, and consistent).

b) Negative causation. Judgments of negative causation seem to be sensitive to what should happen or normally happens. For instance, it seems that it is only because I should have fed my goldfish that we are inclined to say that my omission to do so caused its death. The Queen of England also failed to feed my goldfish, but we do not say that her omission caused its death, and this appears to be because we do not believe that the Queen of England should have fed it. Granting that at least some omissions are causes and that one needs to differentiate causal from non-causal omissions (after all, it was my omission to feed the goldfish, and not the Queen's, that caused its death), a plausible answer to causal overgeneration is to invoke external normative criteria – I was *supposed* to feed the goldfish and the Queen was not. On the other hand, denying the causal status of omissions but allowing them to figure in causal explanations rather than in causal relations (Beebee 2004b) does not constitute an escaping move since one must still distinguish valid from improper causal explanations.

c) Relata individuation. If causation is an extensional relation out in the world, it should hold independently of how relata are described. But there is open conflict on the

appropriate conception of relata that would justify extensionality – on a granularity scale, the options range from the more coarse grained relata (events) to the fine grained (e.g., facts, tropes, states of affairs). If co-referential event nominals cannot be substituted *salva veritate* in a causal claim (for instance, one cannot substitute 'The height of the climber's fall caused his death' with 'The climber's falling *from a height* caused his death', or 'Socrates's *drinking hemlock* at dusk' with 'Socrates's drinking hemlock *at dusk*'), then perhaps extensionality requires reconsideration (Achinstein 1983).

d) Legal causation. Notions such as proximate or intervening causation[16] used by legal theory to explain certain events with the aim of assessing culpability or permissibility have been criticised as being mere doctrinal devices or variable evaluative concepts rather than markers of objective causal connections (Moore 2009). These capricious causal notions are also common currency for the historian, social theorist, or the ordinary person, so a challenge of their respectability in the legal domain should activate at least a partial reconsideration of our current causal intuitions.

e) Experimental evidence. Hitchcock and Knobe (2009) have argued that causal intuitions are determined in part by judgments about the relevance of counterfactuals and relevance is determined by various norms. For instance, if an administrative assistant and a professor take pens from reception and equally cause the department receptionist to be unable to take an important note, the individual who was not supposed to take pens is deemed to be more of a cause. The point is that norm-violations render certain counterfactuals more salient – 'had the assistant not taken pens, the department receptionist wouldn't have been left without any pen' – which in turn affect causal intuitions. In addition, sometimes judgments of causation *just are* judgments of responsibility, i.e., they only mean that an agent violated a norm. Lastly, Alicke, Rose and Bloom (2011) argue that norms *bias* rather than essentially determine causal judgments (they are a mere after-judgment effect). To them, norm-violation determines judgments of causation *via* judgments of blame, which suggests that blaming might actually generate biased causal judgments.

[16] A proximate cause is the nearest, sufficient event setting in motion a train of events that lead to a legally recognizable result. An intervening cause is an event occurring from a new, independent source, and which breaks the direct connection between an action and its legal result.

The list is in no sense inclusive. The evidence is scattered as it appears in the literature under distinct headings – the problem of causal selection (Hesslow 1988, Lewis 1973), the interest relativity of causal explanation (Garfinkel 1981), the context sensitivity of causal language (Swanson 2010), or the factors affecting counterfactual availability (Hitchcock and Knobe 2009). Nonetheless, the general idea pushed by the data is that in many cases causal facts *vary* because they are at least in part determined by a certain perspective rather than being fixed by the world. In general, the presence of normative factors has been noted in the literature, but explained in line with the metaphysical agenda advanced. If those with revisionary inclinations see perspective-dependence as entailing a certain pressure on the shape or substance of theories, the more conservative-minded have the tendency to defend causal realism from conceptual confusion, methodological unilaterality, or other allegations.

Several points need clarification. Some examples (negative causation, relata individuation) concern the metaphysics of causation directly and thus appear to be more damaging to realism. Other examples (legal causation, experimental evidence) seem less detrimental as they might be explained in terms of legal practice or psychological factors. Second, there is an apparent underlying unity of the examples: all presuppose a form of selection (of what counts as *a* cause, a *negative* cause, a *legal* cause, or a causal *sequence*).[17] Third, one does not have to accept all illustrations; what matters is that *some* selections appeal to normative constraints when describing causal facts or explaining causal notions, and are thus incompatible with a realist view.[18] Finally, most

[17] Against the third point, it might be objected that some of the cases mentioned above are not about selection. For instance, it might be argued that negative causation is actually about transcendent entities ('nothings') lacking causal power. Yet this is only one aspect of the problem. At a closer look, offering a solution to causal overgeneration presupposes the selection of relevant from irrelevant negative events – for instance that a particular outcome is due to *my* omission rather than the omission of other persons X, Y, Z.

[18] Taking a moderate line, one does not have to claim that the data tying our intuitive conception to considerations of a perspectival or normative nature counts as decisive evidence against realism, but defend a more modest claim: the presence of this data diminishes the strength of the realist intuition. The realist dismisses the data as contradicting a natural presumption about causation or defines its acceptance as bitter bullet-biting; but those who endorse it may simply aim to question rather than abandon that natural presumption. Such an unpretentious move is harder to counteract. Now perhaps realists are prepared to admit this argument but ask what exactly tempers the realist intuition (and how) rather than repudiate it altogether. I concede that it is difficult to discern what could be proposed in reply. For example, one may defend a response-dependence account of causation, but this would only offer an understanding of how we grasp causal relations rather than affect the core of the realist presumption. Of course, that does not mean that alternative mitigating features could not be brought to light, but until then, the burden of proof remains with the defenders of realism, who need to articulate a plausible picture to explain the ungovernable data.

theories of causation are centred on the nature of connection, whereas perspective-dependence is mostly associated with the nature and designation of relata. Current metaphysics of causation rests on a *separability* assumption – presuming that questions about relation (connection, direction, selection) can be answered independently from questions about relata (immanence, individuation, adicity). Most theories of causation are theories of connection riding on the idea that ontological and conceptual analyses must attempt a reduction to processes, probabilities, or dependence relations (and if reduction fails, then mixed views, primitivism, or eliminativism are straightforward retreat choices). Yet this assumption is not obvious and some philosophers cautioned, if only briefly, that selection of relata is as much part of the meaning of causation as connection is:

> The contrast of cause with mere conditions is an inseparable feature of all causal thinking, and constitutes as much the meaning of causal expressions as the implicit reference to generalizations does. (Hart and Honoré 1985: 12).

> No theory of the causal relation (…) can be the whole story of causation. (…) Any relation needs relata, whether it is intrinsic or not. So the problem of missing relata hits any relational analysis of causation. (Lewis 2004: 282).

If these concerns are serious, separability requires further justification. If the normative aspects present in our intuitive conception of causation are usually associated with the nature and designation of relata and that affects the analysis of connection – e.g., preemption affects probability views (Good 1961, 1962, Lewis 1973, Menzies 1989), disconnection affects process views (Ehring 1986, Hitchcock 1995b, Dowe 2000, Schaffer 2000, 2001), missing relata affect the intrinsicness of causation (Mellor 1995, Lewis 2004) – then a 'whole story' account of causation should observe both intuitions.

1.3 Approaching Intuitions

I have spelled out two incompatible intuitions which I am taking to be central to our conception of causation:

(a) Causation is an objective and mind-independent relation between events located in space-time;

(b) Causation is tied to considerations of a perspectival, projected, or normative nature.

I will next examine two distinct, realist-minded, and representative ways of approaching this incompatibility. The first is a familiar reconciliatory project that safeguards realism by taking causation to be objective and treating perspective-

dependence as an intuition about causal language (Section 2). The second is a utility-based project that takes a version of realism to be supported by considerations of theoretical advantageousness (Section 3). Irrespective of the route taken, I shall argue that defending an unambiguous realist stance about causation is more difficult than expected, as deep-seated ambiguities affect the selection principles involved in both describing and modelling of causal facts. By extension, the theories aspiring to tie outcome responsibility claims to a naturalistic order will need to re-examine their commitments. As causal realism is a multifaceted issue, let me restate three points about my focus:

a) Unless otherwise specified, I shall take causal realism to describe a view about the existence of an objective and mind-independent relation (as mentioned in Section 1.1).

b) I will not examine the relation between causal realism and the scientific image conveyed by contemporary physics (but see Price and Corry [2007] for a detailed analysis of this relation).

c) It is not my intention to add to the arguments against causal realism directly (but see Hume [1739], Russell [1913], Wittgenstein [1922], Putnam [1984], van Fraassen [1980], Blackburn [1993], Spohn [1993], Chakravartty [2005], Beebee [2006], Esfeld [2009]). In contrast, my aim is to show the limits of the *strategies* explaining our incompatible intuitions and how they may bear implications for outcome responsibility attributions.

2. Strategy I: Reconciliation

2.1 Two Contentious Assumptions

A popular reconciliatory move[19] is to place the two conflicting intuitions at different theoretical levels and argue that both can be integrated if we take causation to be objective and causal language to be context sensitive: if the realist intuition fits the metaphysics of causation, the perspective-dependence intuition reflects the dynamics

[19] This is not the only possible reconciliatory move, but the most popular one. One may also explain normative elements by blaming conceptual confusion (Hitchcock 2007a) or methodological unilaterality (Paul 2009).

of our causal talk – hence their apparent incompatibility.[20] The argument goes as follows:

(1) Causation is objective and mind-independent: causal facts are fixed by the world and it alone.
(2) But causation also seems tied to normative determinants: causal facts are at least in part determined by a certain perspective.
(3) A 'whole story' analysis of causation needs to explain both insights and how they fit together.
(4) We can think and theorise about causation in a way that is unaffected by facts about causal language.
(5) When linguistic data is a hindrance to metaphysics, metaphysics takes priority: one should offer a plausible explanation of the linguistic data such that no revisions to metaphysics are required.
(6) The best strategy given (3)-(5) is to argue that while the realist intuition gets the metaphysics right, the perspective-dependence intuition expresses the context sensitivity of our causal talk.

To illustrate, consider Collingwood's example of variation in the description of causal facts:

> [A] car skids while cornering at a certain point, strikes the kerb, and turns turtle. From the car driver's point of view the cause of the accident was cornering too fast, and the lesson is that one must drive more carefully. From the county surveyor's point of view the cause was a defect in the surface or camber of the road, and the lesson is that greater care must be taken to make roads skid-proof. From the motor manufacturer's point of view the cause was defective design in the car, and the lesson is that one must place the centre of gravity lower. (Collingwood 1940: 304).

If the realist-minded will undoubtedly feel that a single complex sequence of events occurred and its causal structure is invariable and objective, those inclined to see causal facts as partly determined by normative determinants will be quick to point out how reports change and disagreement ensues. In reply, a reconciliation theorist would insist that variation only affects what different speakers in different contexts take to be the

[20] Realism/Anti-realism aficionados will immediately recognize this as a version of the argument accommodating metaphysical realism with conceptual relativity (see for instance the 'contextual semantics' account of truth sketched by Horgan and Timmons [2002]).

genuine cause of the accident, but not causal facts themselves. Despite the appeal and actual prominence of this solution, two of its premises are controversial.

Looking at [Premise 4], note how the selective way we individuate particular causal claims is presumed to be independent from the nature of causation itself. Reiterating a Millian theme, Lewis writes:

> We speak of the 'decisive' or 'real' or 'principal' cause[s]. We may select the abnormal or extraordinary causes, or those under human control, or those we deem good or bad, or just those we want to talk about. I have nothing to say about these principles of invidious discrimination. I am concerned with the prior question of what it is to be one of the causes (unselectively speaking). My analysis is meant to capture a broad and nondiscriminatory concept of causation. (Lewis 1973: 558-559).

To Lewis, a legitimate analysis needs to capture a pre-selective notion of causation, untainted by our frequently misleading linguistic practices. But this assumption is contentious. It is not clear that we have such a 'broad and non-discriminatory' concept, on the contrary, all intelligible causal claims featuring in our scientific, moral, or legal discourse are interwoven with selection effects. In different terms, isolating causation from its particularised, selected aspects may render it completely meaningless. Admittedly, one may share Lewis' intuition that such a notion exists at a certain level of abstraction, but it is difficult to imagine what kind of evidence might be brought in its support. As a result, it looks like we cannot really think about causal relations in a way that is completely unaffected by facts about causal talk. If correct, this diagnostic generates a different concern: if the analysis of causation is contaminated by facts about language, and causal language is context sensitive, the metaphysics of causation may end up being underdetermined. This looks like a significant methodological worry. While there might be ways of securing a determinate metaphysics despite the context sensitivity of some terms, causation is notorious in being a *highly* context-sensitive notion (or simultaneously sensitive to more contexts than familiar shifty terms such as 'tall' or 'flat').[21] If this concern is as serious as I think, an obvious consequence is that progress on the metaphysics of causation will be made only in parallel with advances on the conceptual and epistemic fronts. A more troubling consequence is that what exactly determines the metaphysics of causation may remain permanently elusive. But

[21] For instance, token causal claims may be simultaneously sensitive to the context of inquiry (to particular why-questions) and to the context of occurrence (the situation generating a particular claim about what counts as causes and conditions). Regarding the parallel with the context sensitivity as it arises for knowledge ascriptions, Schaffer notes: '[T]he intuitive data for context sensitivity is much stronger and more robust for causal claims, and it includes specific phenomena that have no counterpart with knowledge ascriptions'. (Schaffer 2012: 36).

regardless of the end result, the relevant idea *now* is that placing intuitions at different levels may not get us very far, as one intuition (perspective-dependence) may end up setting constraints on the other (realism). In different terms, while intuitions get explained through a reconciliatory strategy, the metaphysics of causation gets obscured.

Looking at [Premise 5], there is no immediately obvious reason why the first intuition gets the metaphysics right and the second is about the language of causation. An idea could be that we cast the second intuition as one about language because we abide by a priority principle (cf. Swanson 2010):

[Priority] - Try as much as possible to offer a plausible explanation of the linguistic data such that no changes to metaphysics are required (metaphysics takes priority).

But this assumption is also contentious. In light of the previous remarks, **[Priority]** cannot be a clear-cut solution – for instance, taking into account the context sensitivity of our causal talk, at what point do certain changes to metaphysics switch from inappropriate to appropriate? How do we know when a certain linguistic explanation is plausible enough to reflect sound metaphysical distinctions?[22] In reply, one might hold that a reconciliation theorist does not need the priority principle, as her view is supported by inference to the best explanation of the data: it could simply be that this is a good way to capture both of our insights about causation in a single coherent story. However, note that the priority imperative is built into what we take to be the principle of best explanation – inference to the best explanation would be a good way to capture both intuitions just because we are able to cast the shiftiness of causal claims as a mere linguistic matter.

In brief, a reconciliatory strategy assumes that an appropriate analysis needs to capture a pre-selective notion of causation and that metaphysics takes priority, however, both assumptions are contentious.

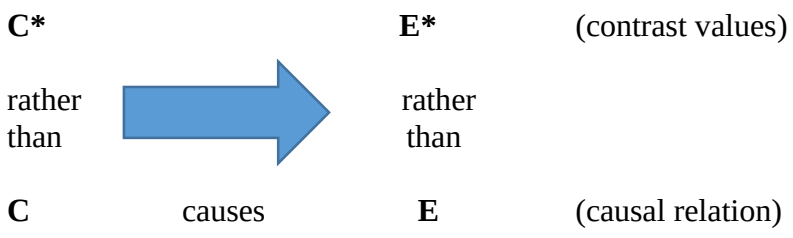
2.2 Contrastivism

Let me take a step back and provisionally set the worries conveyed by [Premise 4] and [Premise 5] aside. If the argument in Section 2.1 is sound, the most popular

[22] The priority principle appears to be based on the deeper desideratum of attaining an equilibrium between what we say and what there is. Yet given the shiftiness of our causal talk it is not clear that an appropriate match with the causal layout of the world is attainable.

implementation of a reconciliatory strategy is a contrastivist theory (Hitchcock 1996, Woodward 2003, Maslen 2004, Schaffer 2005, Northcott 2008). A contrastivist theory of a certain concept holds that the concept in question can be adequately defined only relative to a set of reference frames or alternatives. As an illustration, contrastivism about knowledge holds that whether a knowledge ascription is true in a context depends on the alternatives relevant in that context, e.g., I may know that a bird is a canary rather than a raven, but not know that it is a canary rather than a goldfinch (Schaffer 2008).

Contrastivism about causation treats the data described in Section 1.2 as evidence for the context sensitivity of causal language[23] and takes this evidence to press a requirement on theories to supplement the standard, binary logical form of causal claims with information about their particular context. According to (quaternary) contrastivism, causation is not described with the pattern 'C causes E iff…[condition ϕ]… obtains' – but includes relevant alternatives (or contrast values) tested counterfactually: C rather than C* causes E rather than E* means that C causes E iff, for each of them, had relevant alternatives C* and E* existed and replaced the initial causal relation, C would not have caused E.



For example: 'my forgetting to water the plant caused its death' (of the standard form C caused E) becomes 'my forgetting to water the plant (C) rather than remembering to water it (C*) caused its death (E) rather than its survival (E*)'. Tested counterfactually: 'had I remembered to water it, the plant would have survived' – had C* occurred, E* would have occurred [O(C*)>O(E*)]. Similarly, any token causal claim bears a relation

[23] Causal claims are context sensitive in the sense that the truth or acceptability of causal statements of the form 'C caused E' (or employing similar causal verbs) varies across contexts. It is debatable whether context sensitivity is entirely explained by the semantics of causal claims or by their pragmatic conversational value – whether it is encoded in the meaning of causal statements or reflects the interests and expectations specific to the conversational point. The literature is divided on the issue. A pragmatic view is favoured by Mackie (1980), Lewis (1986a), Bennett (1995), or Swanson (2010), *inter alia*. Semantic context sensitivity is favoured by Hitchcock (1996), Woodward (2003), Maslen (2004), Menzies (2004, 2007), Hall (2007), Schaffer (2005, 2012). Feinberg (1970) hinted at both sources. A brief rejection of (aspects of) the pragmatic view can be found in McGrath (2005) and Menzies (2007). A critical examination of both accounts appears in Schaffer (2012).

to the relevant alternatives or contrast values set by the context in which the token causal claim is made. If there is any variation, it takes place at the level of contrast values C* and E*, but causation 'itself' remains an extensional relation 'out there'. Recall Collingwood's example in Section 2.1, in which something is cited as the cause of the accident depending on the relevant alternatives considered: in a particular context, it is the cornering too fast rather than the defective surface of the road that caused the accident; in a different context, it is the defective design of the car rather than the driver's cornering too fast that caused the accident, and so on.

In more general terms, one may think of contrast values as default events and of causes as deviations from the default events. For instance, consider negative causation: absences or omissions are cited as causes when they violate certain expectations set up by norms, or deviate from the normal course of events. Or consider the difference between causes and background conditions: we take the striking of the match to be the cause of the fire because the presence of oxygen is assigned a default value from which the striking deviates. Or consider legal causation: an agent causes harm because it deviates from the default value which is lawful conduct. Or consider the explanation of how causal relata can be specified in a mind-independent way: contrastivism has no ontological commitment to a fine-grained or coarse-grained conceptions of events; on the contrary, it is a thesis about the structure of causal relations (for an extended discussion of how referential strategies explain the conventional data noted in Section 1.2 – although with distinct metaphysical agendas – see Schaffer [2005] and Menzies [2009]).

Contrastivism is thus able to avoid a revisionary stance, i.e. taking causation to have built-in normative determinants rather than being a purely extensional notion. Drawing a parallel to the context sensitivity of knowledge ascriptions, imagine a version of anti-intellectualism about causation giving priority to perspective-dependence: if anti-intellectualists about knowledge hold that whether a subject knows depends on stakes/interests/etc. (Hawthorne 2004, Stanley 2005, Fantl and McGrath 2009), an anti-intellectualist about causation would argue that whether C causes E depends on practical interests shaped by norms or expectations. Such a view would start from the context sensitivity of causal claims and take the problems of developing contextualist/pragmatic accounts of the variable character of causal claims to support the idea that causation presupposes hidden normative determinants, which in turn would entail a revisionary, anti-realist metaphysics. However, contrastivism denies this

option: causal language can, but causation cannot be dependent on relevant aspects of conversation or elements of social convention. While such an answer may simply showcase the strength of the realist intuition, or may be based on dissimilarities between the knowledge relation and the causal relation, it is worth noting that a 'different levels' approach manages to avoid the unpopular, anti-realist consequences of a revisionary stance.[24]

But I have doubts that contrastivism as an implementation of a reconciliatory strategy succeeds in protecting a sharp realist concept of causation. At a closer look, significant inaccuracies persist. Note first the inconvenient but familiar fact that many causal claims do not make any reference to relevant contrasts. Even if such relativisation were evident – it is usually taken to be implicit because obscured by the surface form of causal talk – it would provide no clear mark of the difference between objective and subjective causal claims. There is simply no discernible sense when and to what extent a token causal claim should be explained in terms of facts about causal language or about causation 'itself'. One needs to be clear about how implicit relativisation to external reference frames does not threaten the objective causal claims endorsed by natural and social sciences. In fact, an illuminating view should not stop at mentioning the implicit presence of actual reference frames, but offer an account of their selection that specifies criteria for successfully fending off arbitrariness. However, referential theories – most notably contrastivism (Schaffer 2005, Northcott 2008) –

[24] Alternatively, one may argue for a description of normative determinants in non-normative terms (McGrath 2005). On such analysis, causation depends on what we take to be the normal or expected, but normality can be analysed in broadly naturalistic terms, e.g. as when we say that it is normal for hearts to pump blood (the normal biological functioning) or for rivers to flow downhill (the normal physical functioning). If we manage to define the normal in non-normative terms, we could accept that causation has a normative aspect (or so the argument goes). However, it is difficult to define a heterogeneous notion like 'the normal' in a unitary fashion. For instance, there is indeed an analogy between norms of functioning in physics and biology, and human conventions – e.g., driving on two sides has the function of avoiding collisions (*simple*), or the legal system represents a convention with the function of achieving justice (*complex*). However, it is only a surface analogy: in fact, human conventions are much more fluctuant. When making overall normative judgments aggregating statistical, prescriptive, and functional senses of normality, there is a tendency to consider the more naturalistic senses on a par with moral or legal 'normality' – which motivates McGrath's analogy. But using statistical and functional regularities as a heuristic for the prescriptive sense will not amount to an equality of status. If they did, it would be much simpler to bridge the factual-normative gap (the way things are and the way they should normally be). In addition, there is a shadow of equivocation in the claim that the natural and the normative are not clearly opposed (e.g., when we say that 'it is normal for the planets in our Solar System to orbit around the Sun' or that 'it is normal for certain animals that are prayed on to have obliterative patterns or colours'). In a sense, the natural and the normative are opposed *ex vi termini*. But it is a completely different thing to say that a natural explanation has become normal given its widespread acceptance or compelling character (e.g., evolution theory). It is only in the latter sense that normative determinants could be described in non-normative terms, for a simple reason: the natural and the normal mean the same thing.

remain silent on the selection of reference frames. In general, as Franklin-Hall (forthcoming-a) remarks, the feeling is that referential views aggregate a set of claims of enough vagueness to make them immune to any potential counterexamples, *viz*. they are centred on a reference variable whose value is left implicit and whose setting is inaccessible.[25]

Furthermore, the data described in Section 1.2 may represent evidence for the context sensitivity of causal claims, but a precise understanding of their shifty character is hard to achieve. First, it is not clear whether context sensitivity is a wholly pragmatic or partly semantic affair (Schaffer 2012). Without getting into details, there are reasons to doubt that pragmatics can handle all context sensitivity. For instance, imagine that a person is stabbed to death at a dinner table and one cites as the cause of death the fact that the table had been laid down for dinner with knives on it. On pragmatic grounds, this is unacceptable because dinner table arrangements are irrelevant in this context. But sometimes the feeling is that such claims are obviously false rather than conversationally misleading: it is false to say that the oxygen in the air of England caused the defeat of the Spanish Armada, or that Socrates's drinking hemlock AT DUSK caused his death. In this sense, pragmatics does not *fully* explain the context sensitivity of causal claims because it only accounts for intuitions of irrelevance (and disregarding relevance does not explain intuitions of falsehood). Second, if one agrees with the recent arguments that 'cause' is *semantically* sensitive to salient contrasts, it is difficult to identify the parameter that tracks the salient contrasts. An option would be to say that contrasts are suggested by the overarching causal question triggered by particular causal statements (taking the general form 'Why X rather than Y?'), but it is hard to see how simple why-questions would provide contrasts for both causes and effects, or how contrasts would connect to the truth conditions of a causal statement (Schaffer 2012: 55-59). It is fair to say that 'we do not yet have a clear understanding of context sensitivity as it arises for causal claims' (Schaffer 2012: 36) and as a result, although contrastivism appears as an appealing and elegant implementation of a reconciliatory strategy, its details are still relatively contentious (or, more clearly, the

[25] If not fixed by rules, reference classes are exposed to disagreement. As Hesslow remarks, 'the reference class is often an unconscious entity, which is formed by a variety of logically irrelevant factors of which experience, norms and educational indoctrination are examples'. (Hesslow 1988: 29).

understanding of perspective dependence as a mere linguistic matter becomes elusive when approaching its specific aspects).[26]

Finally, a more intriguing objection is that referential theories like contrastivism run together causation (an objective, extensional relation) and causal explanation (an intellectual, rational, or intensional relation). The argument is as follows:

(1) To realists, it seems right to say that causal explanation is mind- and description-dependent, but wrong to say that causation is mind- and description-dependent.
(2) Any context change affects explanatory salience independently of whether one offers a pragmatic or semantic account of context sensitivity.
(3) But no such contextual restriction will change what exists, i.e., the domain of causal facts which includes *all* causal facts 'whether or not it is always explanatory or contextually appropriate to refer to some of these facts' (Hall and Paul 2013: 35). Therefore,
(4) Indexing causal claims to reference frames – such as contrast classes – will output a theory of causal explanation (mind- and description-dependent) rather than a theory of causation (mind- and description-independent).

Yet it is not exactly obvious why. An idea would be that contrastivist accounts of causation are modelled after contrastivist theories of causal explanation (like those of van Fraassen [1980], Garfinkel [1981], or Lipton [1990], with empirical support from McGill [1991] or Cheng and Novick [1991]), and theorists expect the latter rather than the former to be referential. Indeed, there is considerably more agreement on contrastivist theories of causal explanation than on causal contrastivism. But things are ambiguous, as it is often unclear whether a statement is read or intended as expressing a causal relation or a causal explanation – for example, causal claims such as 'Frequent flying causes deep vein thrombosis (DVT)' may be read in both keys.

As a result, a successful contrastivist implementation of a reconciliatory strategy would need to offer an account of the selection of reference frames, achieve a clear understanding of context sensitivity as it arises for causal claims, and find a way

[26] Very recently, Blanchard and Schaffer (forthcoming) explored the idea that the contrastivist framework could be supplemented with a psychological story about how people access possible causal and effectual contrasts (how alternatives become available in our cognition). However, it is premature to fully evaluate the success of their suggestion at this stage.

to disentangle causation from causal explanation. Two further remarks are in order: a) the previous critical comments underline the limitations of contrastivism as an implementation of a reconciliatory strategy. They are not meant to target contrastivism as a non-reductive theory revising the binary structure of causal claims to handle omissions, preemption, transitivity, selection, or other well-known difficulties of causation; b) even if the mentioned deficiencies of contrastivism are ultimately dealt with, clarification is still required for the first level controversial assumptions (premises 4 and 5).

2.3 An Ambiguity

But even granting an answer to the first two objections – about selecting reference frames and achieving a clear understanding of context sensitivity as it arises for causal claims – I find it difficult to see how contrastivism, or referential theories in general, will manage to disentangle causation from causal explanation, as it is unclear when and whether a token causal claim is read as expressing a causal relation or a causal explanation.

To make sense of this problem, we should start from the idea that any particular caused event stands at the end of a complex causal history. To describe the ontological reality underpinning this multiplicity of causal determinants, the image most often appealed to is that of a sophisticated network.[27] The concept of a causal network is essentially Millian and derives from modelling the generation of effects after paradigmatic mechanical phenomena in which the transfer of motion is regulated by the principle of composition of forces. Along the same lines, effects are the result of particular combinations of two or more causes which add up by a sort of vector algebra – what Mill (1843/1947: VI) calls the principle of Composition of Causes.[28] The resulting network or Millian model is a standard metaphysical representation of the most fundamental level of reality, *viz.* a seamless structure of causal interactions that may extend *ad infinitum*, it is intricate and dense, and may take different interpretations

[27] According to Steward, '[t]he idea of a network is an almost irresistible resource in the attempt to visualise and make manageable our understanding of the highly complicated causal relationships which seem to be revealed by the wealth of alternative, equally true causal explanations we can offer of any event or circumstance'. (Steward 1997: 127).

[28] Beside 'mechanical composition', Mill also talks about 'chemical composition', in which causes combine to generate an effect with emergent properties (properties that cannot be predicted from the properties of the initial chemical substance-like causes) and thus breach the more general principle of Composition of Causes.

depending on the ontology of the nodes (events, facts, tropes etc.) and the character of relations (deterministic or probabilistic).[29]

The question then becomes: how do we get from this great blooming, buzzing confusion of micro-level causal interactions – to use William James's (1890/1981) phrase – to the discovery of singular causal relations, the identification of elaborate mechanisms, or the determination of direct, contributing, total, proportional, root, negative, triggering, predisposing, proximate, or other kinds of 'interesting' causes? How do we break the causal wholeness of the world? One tantalizing speculation is that many of these causes result from a set of restrictive operations applied to the network model – call them *selection constraints*. The idea of selection constraints becomes relevant with Mill's (1843/1947: III, v, 3) observation that in theory any correct assessment of token causal claims would imply stating the sum of all antecedent and adjacent network nodes. Although Mill aims to expose the arbitrariness of causal selection (the way we prioritise causes over background conditions), the more general question concerns the partitioning of causal space – as accepting Mill's challenge would undermine not only the idea of discovering single causes, but our right to use notions such as 'the successor', 'the predecessor', or claim that anything 'brings about' anything else. Here is a brief inventory of possible restrictive operations:

(1) Spatio-temporal constraints. The Millian model operates with the assumption that the causal order shares the same direction as the arrow of time. As a result, individual causal relations describe a transition from particular initial states to target effects – although the transition between determined states is in theory independent of the temporal direction. These initial states are distinct from and do not go as deep as the initial state of the model (which coincides with the earliest state of the universe). Combined with an analogous spatial dimension, we get a sort of spatio-temporal markers, necessary for achieving causal focus and avoiding regression. Accordingly, stating the cause of a certain occurrence means tracing or specifying the process from the spatio-temporal marker(s) to the target effect.

[29] The network model is implemented in neuron diagrams (Lewis 1973) or directed acyclic graphs (Spirtes, Glymour, and Scheines 1993, Pearl 2000).

(2) Logical constraints. Another important restriction is that causes must be sufficient for their effects. For instance, looking for the causes of Alan Turing's death in 1954, it would not do to say that 'Alan Turing's eating an apple caused his death', as most people eating apples do not die. A more adequate reason is that 'Cyanide ingestion caused Alan Turing's death' and the apple was the means by which the fatal dose was ingested. Sufficiency must not be confused with the requirement of *proportionality* (Garfinkel 1981, Yablo 1992) – the idea that causes need to be described in adequate detail[30] – in the sense that less detail entails flouting sufficiency (although both seem guided by considerations of relevance and sometimes overlap).

(3) The causal field (Anderson 1938, Mackie 1955, 1965, 1980). A causal field is a restrictive assumption or set of assumptions made about singular causal statements with the purpose of capturing the specific circumstances of their occurrence. In claiming that X causes Y we assume a determined field and look within that field for the differentia marking off the occurrence of Y from the non-occurrence of Y. For instance, an inquiry into the causes of anger is indeterminate and would need to be further specified as 'What causes anger *in this person*?', or against different or wider causal fields like *rats*, *human beings*, or *animals* in general. Causal sub-regions may compete in providing the cause of an event, and for that reason the choice of a causal field, which is often indeterminate or taken for granted, needs to be made clear. Because it divides the network model in sub-regions and limits what is included in the cause of a particular event, this restriction represents a significant advance in causal mapping. Through its lens, singular causal statements of the form 'X causes Y' are elliptical and need to be expanded into 'X causes Y relative to a causal field F'.[31]

[30] For instance, when looking for the cause of an injury, being hit by a yellow car conveys too much detail, being hit is not enough, and being hit by a car is just right. However, what makes for the appropriate specificity depends on pragmatic rules.

[31] Three remarks:

(i) The notion of causal field assumes causation to be a difference making relation and invites a referential treatment of causal claims.

(ii) Note the presence of two cognitive operations on the Millian model: a) determining a causal field; b) projecting an alternative course of events as a reference value. When sub-regions compete, the causal field is taken to be the source of contrast alternatives because it is identified with a certain conversational context, but the sets of alternatives may be distinct from the causal field.

(iii) The causal field is a vague and undertheorised notion. Causal claims may assume different causal fields in the description of an event – e.g. 'Cyanide ingestion caused Alan Turing's death' vs. 'Alan Turing's birth caused his death' – but there is no clear psychological or linguistic explanation why. The causal field may be a notion pragmatically assumed for conversational purposes or for fixing the truth of causal claims: '[W]hereas Mackie takes the causal field to be a set of assumed conditions pragmatically

(4) Abstraction. The identity conditions of causes (stable entities, events, etc.) or determined causal sequences are specified by abstracting from the seamless structure of causal interactions. For instance, concrete individual objects, dynamic states, or regularities – e.g., trees, storms, planetary movements – are not discovered as such, but abstracted from the more fundamental causal structure described by the network model. Abstraction is a relevance-guided operation of reducing the vast and refined causal information in the model, and it takes various forms – from abstract particulars and basic abstractions to experiential, semantic, or categorical types. As a result, abstraction understood as a high-level specification of aggregative or emergent properties implies a commitment to multiple levels of description.

The list may include other restrictive procedures or refined versions of the mentioned ones. Were these restrictions absent, causation would be – as van Fraassen (1980: 124) warned – 'whatever structure of relations science describes.' But thanks to them we are able to answer questions about causes that conflict with, add to, replace, overlap, displace, presuppose, or complement other causes. Furthermore, these restrictions can be integrated: for instance, certain types of abstraction may use spatio-temporal constraints, causal fields or self-contained causal sub-regions are themselves relevance-determined abstractions, and sufficiency may be a requirement within a causal field – that is, causal fields may function as frameworks for identifying genuinely sufficient conditions. Lastly, selection constraints seem to have two important roles:

(i) Intelligibility: we use selection constraints to make sense of specific causal sequences in the network. Restrictions emerge where our limited cognitive abilities encounter the overwhelming complexity of the model; therefore any description of particular causal facts (events or circumstances) or instance of token causation presuppose them.[32] On this view, selection constraints represent a form of apprehension (causal frames allow us to understand certain things about changes or dependencies in the world).

superimposed on a pre-selective notion of connection, the quaternary theorist relocates the field to determining the semantic value of the causal alternative and the effectual difference.' (Schaffer 2007).

[32] Although there might be an intermediary level of uninformative selection (or diminished-intelligibility selection): 'X causes Y' could be described as e.g., Node8934 causes Node9045 in the Millian network.

(ii) Explanation: we use selection constraints to substantiate causal explanations. More precisely, sufficiency and spatio-temporal restrictions are used to establish process-explanations describing the transition from an initial state to a target effect (Salmon 1984, 1994, 1997, Rueger 2006), while causal fields or similar abstracted sub-regions provide the basis for accounts of contrastive explanation (van Fraassen 1980, Garfinkel 1981, Hesslow 1983, 1988, Lewis 1986b, Lipton 1990, 1991, Barnes 1994, Hitchcock 1996, 1999, Schweder 1999, Ylikoski 2007).[33]

In short, causal space may be partitioned for intelligibility or explanatory reasons. The two functions are closely related and often run together, which contributes to the generation of an essential ambiguity in the delimitation and description of causal facts. 'X causes Y' is by definition ambiguous between a way of apprehending specific worldly relations and providing particular explanations. This ambiguity needs to be distinguished from Shorter's (1965) and Davidson's (1980a: 162) suggestion that 'causes' is semantically ambiguous between the 'real' relation of causation and the sentential connective characteristic to causal explanations.[34] It is neither the term 'cause' that introduces the ambiguity – indeed, 'cause' does not seem to be ambiguous in the way that homonyms like 'bat' or 'bank' are ambiguous – nor the nature of relata, that is, it is not that events as relata indicate causation, while facts as relata indicate causal explanation. On the contrary, it is the function of selection constraints – or the nature of causal selection more generally – that invites ambiguities. In most cases, it is simply not transparent when such restrictions are meant to describe genuine causes or causal sequences, and when are employed to offer causal explanations. The equivocal character of selection constraints inevitably leads to ambiguities about causation, namely:

a) to inconsistent intuitions about its nature: it seems that we tend to be causal realists if our selection of token causal sequences is intelligibility-based, and are inclined towards anti-realism if our selected perspective is explanation-based.

[33] For an excellent discussion of process- (or *how*-explanations), contrastive- (or *why*-explanations), and their relation, see Botterill (2010).

[34] Davidson considers claims such as 'The failure of the sprinkling system caused the fire' or '[T]he collapse was caused, not by the fact that the bolt gave way, but by the fact that it gave way so suddenly and unexpectedly' and argues that the 'caused' of such sentences is not the 'caused' of causal relations, but is best expressed by 'causally explains'. In this sense, Davidson clarifies, 'caused' is 'sometimes a relation, sometimes a connective'. (Davidson 1980a: 162).

b) to inconsistent intuitions about the analytical priority of causation and causal explanation: for instance, if Salmon (1984), Lewis (1986a), and Kim (1981, 1993d) analyse explanation in terms of causation, Hanson (1958), Ruddick (1968), and Scriven (1975) analyse causation in terms of explanation.

2.4 Reconciliation and Responsibility Assessment

Let me take a step back and review the reconciliation strategy and its problems. The idea behind a reconciliatory strategy was to place our conflicting intuitions about causation at different theoretical levels and argue that both can be integrated if we take causation to be objective and causal language to be context sensitive: if the realist intuition fits the metaphysics of causation, the perspective-dependence intuition reflects the dynamics of our causal talk. I pointed out two controversial premises of the argument supporting this strategy: the idea that an appropriate analysis needs to capture a pre-selective notion of causation, and the idea that metaphysics takes priority. Both premises require further theoretical attention. However, I left them aside for a moment to examine the most popular implementation of a reconciliatory strategy, i.e. contrastivism about causation. Sadly, it is not very clear that causal contrastivism is successful in safeguarding an objective concept of causation. Beside other problems, contrastivism appears to run together causation (an objective, extensional relation) and causal explanation (an intellectual, intensional relation), which indicates a deeper ambiguity in the description of causal facts, *viz*., it is unclear when and whether a token causal claim is read as expressing a causal sequence or a causal explanation.

Now how do these considerations inform moral assessment? As a general point, any pressure on the objective and mind-independent character of causal sequences will affect the views which hope to tie attributions of outcome responsibility to an objective relation. Accordingly, any doubts cast over reconciliation as a successful strategy – successful in its project of safeguarding the objectivity of causation by claiming that perspective-dependence intuitions reflect the dynamics of our causal talk – should ring alarm bells for a moral theorist with naturalistic ambitions such as Moore (2009).

As a particular point about the ambiguity in the description of causal facts, call *responsibility realism* the following position: an agent A is morally responsible for an outcome O in virtue of the fact that A bears to O an objective, determinate relation R, and R is the causal relation. Call *responsibility irrealism* the following view: the fact that an agent A is morally responsible for an outcome O is not grounded in any

objective, determinate relation R. Rather, the relation obtaining between agent A and outcome O is theoretical – say, the presence of an outcome O is *explained* by A's being responsible for it. At an initial stage, it does not seem to make much difference whether responsibility assessment is based on causation or causal explanation: it does not seem to matter if an outcome is caused by an agent or causally explained by some of her relevant features. As a result, the ambiguity in the description of causal facts is not in principle troubling for the moral philosopher interested in the grounds of responsibility. However, at a second stage it matters if explanation itself is grounded or not in an objective relation. As Kim (1993d: 229-230) notes, a realist about explanation would argue that C is an explanans for an explanandum E if C and E are underpinned by real events *c* and *e*, and if *c* and *e* stand in an objective, determinate relation (the causal relation). In simpler terms, C explains E in virtue of the fact that *c* causes *e*. Alternatively, a critic of explanatory realism would see explanation as a matter of logical, conceptual, or epistemic relationships amongst items of knowledge rather than being grounded in an objective relation. It therefore follows that ambiguity is detrimental to responsibility attributions as long as one subscribes to explanatory irrealism and severs the tie between explanation and causation. In different terms, as long as ambiguity persists and explanation is taken to be an intellectual, rational or intensional relation, moral assessment will not be provided with a naturalistic basis.

2.5 So Far

Strategy I addresses our incompatible intuitions about causation through a reconciliatory project that safeguards causal realism and treats perspective-dependence as an intuition about causal language. But to be successful, a reconciliatory project still needs to clarify important first level assumptions about language and metaphysics, and explain the selection of reference frames, the implementation of context sensitivity as it arises for causal claims, and the apparent running together of causation and causal explanation. In addition, the selection constraints involved in the description of causal facts play an ambiguous role: they may be devices for understanding instances of an objective and mind-independent relation, but also context-related or value-guided devices for causal mapping with explanatory aims. The problems emerging at the interface between causal language and the metaphysics of causation render realist sympathies difficult to stand by. And the essentially equivocal character of most relevant causal claims may deprive (outcome) moral responsibility of a naturalistic

basis. A further remark: the presence of selection constraints shows that, after all, a basic form of referentialism is correct: 'X causes Y' means X causes Y given certain constraints combining to create a referential point in relation to which we can make sense of that particular causal claim. In this sense, intelligible causation is real but relative.[35] But in a different sense, all relativisation strategies specific to referential theories are problematic, so the verdict on the right standpoint is still out. As a result, advocating an unambiguously realist conception of causation is not as easy as one might have initially thought.

3. Strategy II: Utility

3.1 A Different Angle

A different approach is to favour a project promoting considerations of theoretical utility. This project is centred on a concept of causation that is objective in the sense that it offers an adequate and non-arbitrary encoding of the dependence relations amongst phenomena. This notion is the focus of the recent literature on causal modelling (e.g., Spirtes, Glymour, and Scheines 1993, Pearl 2000, Halpern and Hitchcock 2010, 2013) and it is theoretically useful in the sense that causal models generate systematic and accurate predictions about the physical or social systems they represent, i.e. they allow the empirical testing of causal hypotheses in the light of statistical data. Particular to Strategy II is its oblique line of attack: indeed, we have incompatible intuitions, but if no near-at-hand reconciliatory solution is available, those of a realist persuasion might focus on the causal modelling framework as the best theory out there delivering a scientific, hence non-arbitrary, conception of causation. As Halpern and Hitchcock note,

> [Q]uestions about causality are best addressed by dealing with the methodological problem of constructing a model that correctly describes the effects of interventions in a way that is not misleading and ambiguous. (Halpern and Hitchcock 2010: 396-397).

The structural equations specific to causal models express effects of *interventions* – e.g., what happens with a window if hit by a rock, what happens with a migraine if treated with a pain reliever, and so on – and as such they describe real and objective

[35] Further work could be done to elucidate the psychological underpinnings of selection constraints and their application criteria – attention, discrimination, comparison, or pattern-perception.

relations in the world. However, the choice of variables, their values, and the selection amongst competing but equally apt causal models are, to some extent, subjective (Halpern and Pearl 2005, Hall 2007, Hitchcock 2007b, Halpern and Hitchcock 2010). It is up to us what to include in a model, as 'nature does not provide a uniquely correct set of variables'. (Halpern and Hitchcock 2010: 394). Given that even relatively non-controversial systems can be modelled with different aspects in mind and thus provide incompatible verdicts about actual causation,[36] one needs to justify the modelling choices made. Loosely speaking, justification in this context means offering objective reasons for the adequate character of the modelling choices made, explaining why the model is an apt one, and how we know it is an apt one. A causal model is apt or appropriate if it offers a correct encoding of the dependence relations amongst phenomena and leads to accurate and complete predictions and interventions. The problem of justification has at least four aspects.

(i) One aspect concerns the *number* of variables included in a model: is there a rule limiting how many variables could be included in a particular model? Is a restricted set of variables a virtue or a disadvantage of models?

(ii) Another aspect concerns the *values* of variables. Halpern and Hitchcock (2010: 394) define a variable X' as a refinement of X if for each value x in the range of X there is a subset S of the range of X' such that X = x just in case X' is in S. Is there a moment when it is preferable or opportune to replace a variable with a refinement? And what about the *relations* amongst variables in a chosen set – for instance, can the choice to include variable X in a model be accompanied by a requirement to include a different variable Y and/or exclude another variable Z?

(iii) It has been recently argued that structural equations accounts need to be supplemented with a distinction between the default and deviant states of a segment of reality (Menzies 2004, Hall 2007, Hitchcock 2007, Halpern 2008). If added to standard causal models, default information also requires justification.

(iv) Lastly, given the plausible assumption that a determinate system or phenomenon can be aptly modelled in many different ways (or that many causal models will pass the tests for aptness), modelling options need to be justified.

[36] For instance, modelling the actual cause of a plane crash, a flight instructor may point out to the pilot's error, a meteorologist may mention the weather conditions, a sociologist may indicate deficient managerial practices and employment relations, a psychologist may emphasize a stressful episode in the pilot's mental history, and so on. Since all are reasonable answers, a natural question is what counts as an appropriate choice of variables – or more generally, what makes a model an appropriate model.

As Paul and Hall noted, '[i]t is an excellent question, inadequately addressed in the literature, precisely what principles should guide the construction of a causal model.' (Paul and Hall 2013: 18-19). Although a good deal can be said to reduce the arbitrariness in the construction and selection of appropriate causal models (e.g., Halpern and Hitchcock [2010, 2013], Halpern [2013]), I shall argue that considerably more work is required to obtain systems of equations accurate and complete in their predictions and interventions. Amongst the reasons immediately coming to mind, one may note that although the conception of causation envisaged is scientific, modelling is an 'art'; although one hopes for objective constraints on model aptness, certain constraints are strikingly vague;[37] and even if one hopes that appropriate models tend to agree in clear-cut cases, there are important lessons to be learned from special examples. But prior to examining the problems of justification, let me briefly introduce causal models.

3.2 Causal Models

Following Pearl (2000) and Halpern and Pearl (2005), a causal model consists of a system of chosen variables, their allocated values, and a set of structural equations capturing the dependence/influence relations amongst variables. For instance, when trying to determine whether a plane crash was caused by human error or faulty navigation systems, the situation may be described by three variables:

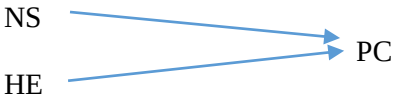
PC for the plane crash, where $PC = \{1\}$ if there is a plane crash and $PC = \{0\}$ otherwise.
HE for human error, where $HE = \{1\}$ if a human error occurred and $HE = \{0\}$ otherwise.
NS for navigation systems, where $NS = \{1\}$ if the navigation systems were faulty and $NS = \{0\}$ otherwise.

If the plane crash is due to either human error or faulty navigation systems is modelled by the equation $PC = \max(HE, NS)$. If the plane crash requires both causes, the situation is modelled as $PC = \min(HE, NS)$. Variables are *exogenous* if determined by factors outside the model and *endogenous* if ultimately determined by exogenous variables. For instance, *PC*, *HE*, and *NS* are endogenous, while factors we usually take for granted

[37] For instance, the condition specified in Hitchcock (2007b: 503): The variables should represent *enough* events to capture the *essential structure* of the situation being modelled – where notions such as enough variables or essential structure are vague or context-sensitive.

– such as gravitation, friction, or the pilot's training history – are all determined by exogenous variables.

More formally, a causal model *M* is a pair (*S, F*). *S* is a signature which describes a certain system and lists its variables and their values. *F* is a set of structural equations connecting the values of variables. A signature is a triple (*U, V, R*), where *U* is a set of exogenous variables, *V* is a set of endogenous variables, and *R* assigns to every variable $Y \in U \cup V$ a non-empty set *R* (*Y*) of possible values (i.e., the *range* of *Y*). The set of structural equations *F* assigns to each endogenous variable $X \in V$ a function *Fx* such that $Fx : (xu \in UR(U) \times (xy \in V-\{x\}R(Y)) \rightarrow R(X)$, which means that *Fx* determines the value of X, given the values of all other variables in $U \cup V$. For instance, modelling the plane crash involves a straightforward signature $S = \langle U1=\{HE\}, U2=\{NS\}, V=\{PC\}, R\rangle$, where *R* maps either *HE* = {0,1} or *NS* = {0,1} (or both) and maps *PC* to {0,1}. The set of structural equations *F* describes the natural dynamics of the signature such that $HE = \{0\} \& NS = \{0\} \rightarrow PC \{0\}$, $HE = \{1\} \& NS = \{0\} \rightarrow PC \{1\}$, $HE = \{0\} \& NS = \{1\} \rightarrow PC \{1\}$, $HE = \{1\} \& NS = \{1\} \rightarrow PC \{1\}$, or $HE = \{1\} \vee NS = \{1\} \rightarrow PC \{1\}$, e.g. This array of structural possibilities is distinguished from what actually happens, *viz.* $HE = \{1\} \& NS = \{0\} \rightarrow PC \{1\}$, e.g. Finally, associated with every model is a directed acyclic graph[38] capturing causal structure:



If *causal structure* is an objective feature of the situation described, the choice of variables is, to some extent, subjective and underlies particular claims of *actual causation* – e.g., 'The search for warmer grounds cause the autumn migration of Canada geese'. Actual causation refers to token sequences of events holding in appropriate circumstances (i.e. with certain facts/variables held fixed or screened off). The formal characterisation of actual causation is fairly complex (for a detailed account, see Halpern and Pearl [2005], Halpern and Hitchcock [2010], Halpern [2013]); however, a straightforward rendering relies on the notions of *dependence* and *directed path* – there is a directed path in M from endogenous variables X1 to Xn if the value of Xi+1 depends on the value of Xi (for $1 \leq i < n$) – and describes actual causation as follows:

[38] In directed acyclic graphs, causal structure is represented by an acyclic Bayesian network, i.e. there is no cycle V1, …, Vn, V1 of endogenous variables forming a directed path from V1 back to itself.

X=x is an actual cause of Y=y in the model M iff the following three conditions hold:
(i) Actual values of X and Y are x and y (i.e., X and Y are actual events).
(ii) There is a directed path DP from X to Y and a value configuration *w* of the set of variables W outside the directed path such that:

a) Had W taken the values w, the variables on the directed path DP would have still taken their actual values.
b) Had W taken values w and X value x, Y would have taken value y.
c) Had W taken values w and X some other value x', Y would have taken some other value y' [in a regular/stable way] (Hitchcock 2001, 2010).

In simpler terms, directed paths are chains of unaffected dependence or derivation from X to Y given a specified setting of the background variables W (where W are off path variables or background nodes), such that any change in the value of X entails a regular change in the value of Y. For instance, HE is the actual cause of PC if HE = {1} and PC = {1}, the off path variable NS = {0}, and given that NS = {0}, if HE = {0} then PC = {0} (or *HE* = {0} & *NS* = {0} → PC {0}). Although far from being complete or problem-free, this account manages to convey the key insight behind the causal modelling framework, *viz.* that causation is correlation under intervention: X causes Y if intervening on X is a way of intervening on Y (given a value configuration w of off path variables W).[39] Having defined what causal models are, I will next discuss the question of justification.

3.3 Can We Justify the Number of Variables in a Model?

Halpern and Hitchcock (2010) note that adding more variables to a model affects causal verdicts. If we had an infinite series of models M_1, M_2 … with the variables in M_i being X0, …Xi+1, Y, and $M^{i+1}_{Xi+1 \leftarrow 1} = M_i$ (where M^{i+1} can be understood as an extension of M^i), it is possible that whether X0 = {1} is a cause of Y = {1} may vary across the range of models. There is, therefore, a requirement of *stability* on apt models, *viz*. they should conserve causal verdicts when supplemented with further variables. For instance, one should be able in principle to argue against using a model M_4 to show that X0 = {1} is a cause of Y = {1}. A series of models M_1, M_2 … stabilizes if there is a *k*

[39] Naturally, intervention is only a way – more accurate in character – to describe what is for X to cause Y given certain background circumstances. One may identify causal relations simply by inference from non-experimental observations, but they are more exposed to the status of claims about mere correlations.

such that for all $i \geq k$, M^i delivers the same verdict – and thus provides a reason to claim that M^k is a sufficient model.

But *Stability* is only a rule of thumb, as defining the number of 'enough' variables in terms of stability and stability in terms of sufficiency sounds rather vague. In addition, *Stability* seems to imply that we are already aware of the causal verdict we aim to conserve (such that no other added variables will change it), however, simple relations of causal dependence may be confirmed or overturned once the model is extended and context added into the picture. Moreover, given that a modeller may be aware only of a certain spectrum of variables and her partial choices may fail to capture all the relevant details of a situation, adding more variables should somehow extend the knowledge we already have. But at the same time, experience shows that there is almost always another (causal) story to be told. *Stability* should in some way strike in between, and that makes it a hardly conclusive constraint. In different terms: a model contains enough variables when it represents facts accurately, and it is accurate when it delivers the right causal verdicts. However, a) if it delivers the right causal verdict, it is hard to say why the number of variables is 'not enough', and b) we are often not aware of the right causal verdicts so we could offer a correct representation of facts with a suitable number of variables (obviously, a model cannot contain 'enough' variables when delivering a causal verdict and 'not enough' or 'too many' when delivering another).

A potential guideline is that the addition of more variables is sensitive to the 'topology' of models (Halpern and Hitchcock 2010). In the plane crash example, two directed paths run into the variable PC, from human error (HE) or from faulty navigation systems (NS), and any interpolated variable along those paths will not alter causal structure (e.g., we might add variables such as PR for premature flap retraction, AS for the aircraft stalling, AL for the subsequent altitude loss etc.). In a case of causal preemption, however, things are different. Suzy and Billy throw rocks at a bottle; Susan's rock arrives first and shatters the bottle, but had she not thrown, Bill's rock would have shattered it a split second later. Initially, we get a similar three-variable model:

BS for the bottle shattering, where *BS* = {1} if the bottle shatters and *BS* = {0} otherwise.

ST for Suzy's throw, where *ST* = {1} if Suzy throws and *ST* = {0} otherwise.

BT for Billy's throw, where *BT* = {1} if Billy throws and *BT* = {0} otherwise.

Nonetheless, since the model fails to capture the distinction between Suzy's rock hitting the bottle first and both rocks hitting it simultaneously, one needs to add two further variables, *BH* for Billy's rock hitting the bottle, where *BH* = {1} if it does and *BH* = {0} otherwise, and *SH* for Suzy's rock hitting the bottle, where *SH* = {1} if it does and *SH* = {0} otherwise. The 'topological' difference is that the additional variables *BH* and *SH* create an additional directed path rather than simply being interpolated along the previous paths, so adding further variables will affect verdicts of actual causation only if their addition alters the topology of the model (Halpern and Hitchcock 2010: 395).[40]

Yet unless more detailed modelling is desired, there is no immediate reason to interpolate new variables along an established causal structure. Does it follow that any meaningful addition of variables will in principle bring a change in topology or lead to a level of complexity that may compromise stability? Hard to say. In addition, actual causation verdicts seem to be affected not only when additional variables create a topological difference: it seems that a simple change in focus is enough to capsize causal results. To illustrate, consider the example of sex bias in postgraduate admissions discussed by Bickel, Hammel, and O'Connell (1975). When considering the data from the university as whole, the results showed that being a male was a cause of being admitted (a negative bias toward the admission of females). However, when considering the data department by department, the results showed a minor negative bias toward the admission of males. The conflict in the data emerged from the large number of female applications to departments with a high rate of rejections (independent of the sex of applicants). Tweaking the example further, one may introduce a variable for the level of competence (e.g., checking the number of applications from Ivy League universities) and discover again a negative bias toward the admission of females. The point is not that adding more variables may overturn causal results through a change of focus along (roughly) the same directed path; rather, as remarked by Suppes (1984), it looks like there is no 'right' stopping point in the refinement of data and therefore no end to its analysis in a practical sense. Moreover, evaluating the data immediately after collection, it is easy to come up with different sorts of hypotheses and artificially create events and define interventions that would

[40] An 'additional directed path' must be taken in the sense of generating a distinct causal verdict rather than in the sense of something over and above individual throws ST and BT.

invalidate any prima facie cause – which brings us back to more fundamental questions like 'What is a serious hypothesis?'[41] or 'How can we tell when a model misrepresents the facts?' The absence of a well-defined stopping point in the refinement of data confirms the intuition that there is no unique way of constructing a model and therefore weak hopes for offering 'a more complete mathematical characterisation of the conditions under which the verdicts of actual causality remain stable under the addition of further variables' (Halpern and Hitchcock 2010: 395).

3.4 Can We Justify the Values of the Variables Included in a Model?

The values of variables also need to be chosen appropriately. To capture the details of a situation, a variable may take multiple values – e.g. in 'Saying hello loudly causes my neighbour to be startled' (Goldman 1970), variable *H* takes value {1} for saying hello, *H* = {0} for not saying hello, *H* = {2} for saying hello loudly, and *H* = {3} for saying hello softly. If it is a disputed metaphysical question whether values {2} and {3} express different fine-grained events *H1*, *H2*, or levels of detail of an event *H*, the metaphysics of event individuation is not relevant to causal modelling. Under the assumption that the world is described in terms of variables and their values, what matters is the correct description of interventions and not the metaphysical nature of events.[42]

One of the concerns is that if models describe dependence relations, certain variables come with implicit presuppositions that may constrain the values of other variables (Halpern and Hitchcock 2010). For example, when modelling Caesar's death (*D* = {1}) by considering a directed path with a remote variable *CB* = {1} for Caesar's birth and a proximate variable *BS* = {1} for Brutus's stabbing (with *D*, *CB*, and *BS* taking value {0} otherwise), it is easy to remark that *CB* = {0} renders variable *BS* useless for any of its values. In contrast, any value taken by *BS* would imply that *CB* = {1}. As a result, if the modeller chooses to include variable *BS* in the model and *BS* =

[41] One the one hand, a condition introduced by Hitchcock (2001) on the aptness of models is that we should not consider the scenarios we are not willing to take seriously. However, looking at the data that does not fit favorite hypotheses or desired theories is one of the best ways to arrive at new hypotheses and theories.

[42] Although it remains unclear how to understand the relation between variables and their values: is it the relation between type causation and token causation (where type causation relates variables and token causation relates values of variables)? Or it is just that values of variables express aspects of token causation? And what is a variable, after all? Are the properties or characteristics of a system represented by variables, or by values of variables?

{0,1} → *CB* = {1}, she cannot conclude that Caesar's birth is a cause of his death, as the alternative *CB* = {0} is not contemplated.

Does this choice misrepresent causal facts? At a closer look, if variable *CB* = {1}, variable *BS* would count as interpolated along the directed path linking Caesar's birth to his death and in principle it should not affect a causal verdict like *CB* = {1} → *D* = {1}. In reply, *D* = {1} might be taken to be a logical rather than causal consequence of *CB* = {1}, but it describes something different from *D* = {1} if *BS* = {1} were the case. On the other hand, it might be argued that it is counterintuitive to count someone's birth as a cause of his death. Yet this claim is not completely implausible. To illustrate, consider José Saramago's alternative story of Lazarus resurrection, when at the very last moment Mary Magdalene places a hand on Jesus' shoulder and says that no one has committed so much sin in his life that he deserves to die twice.[43] In the novel, the act of restoring Lazarus to life – his rebirth in this context – would have caused him to die for a second time (and it is precisely this connection that Mary's contention is based on). Pushing on a different front, it may be argued that we are interested in what made the difference between Caesar's death and his survival in a particular context of his life, and for that reason the presence of variable *BS* in the model would relegate *CB* = {1} at the status of a 'background variable'. However, this move reiterates the problem of causal selection at the level of variables. An option would be to deal with it by introducing referential values (such as contrasts or defaults), but again, referential information also requires justification. Introducing a variable X in a model may therefore entail the inclusion or exclusion of another variable Y, so it is an important question when it is desirable or suitable to do that.

3.5 Can We Justify Default Information?

The excitement over the structural equations framework and its success in pinning down the causal structure of any situation has been toned down by its apparently inappropriate answers in cases of negative causation or cases with isomorphic structures but different causal stories, e.g. – which apparently show that determining actual causation involves more than structural equations (Hall 2007, Hitchcock 2007b,

[43] 'This being so, and with everything necessary in place, the power and the will to use that power, all Jesus has to do is stretch out his arms to that body abandoned by its soul, and say, Lazarus, arise, and Lazarus will rise from the dead, because it is the will of God, but at the very last moment Mary Magdalene placed a hand on Jesus' shoulder and said, No one has committed so much sin in his life that he deserves to die twice, and dropping his arms, Jesus went outside to weep'. (Saramago 1994: 361-362).

Halpern 2008). A recently considered solution is to supplement the usual formalism of causal models with a distinction between *default* and *deviant* states of a system (Hall 2007). In very general terms, defaults represent what is *normal*, *expected* or *taken for granted* with respect to the states or evolution of a system; anything else is a deviation. We judge something to be a cause of a deviation in a system if it has made a difference to the default state or behaviour of that system. For example: instead of moving away according to Hubble's law (default), our galaxy moves towards and will collide with Andromeda galaxy in approximately 4 billion years (deviation), because forced by the gravity of the dark matter surrounding both galaxies and resisting expansion (cause).[44] Similarly, 'The presence of oxygen caused the fire in the shuttle Discovery's rear engine compartment' makes sense only if precautions against oxygen leaks are normally taken, that is, if the default state excludes the presence of oxygen (or takes its presence to be unexpected or abnormal). Thus adding information about what it is normal, expected or taken for granted with respect to the states or evolution of a system helps us determine what (actually) caused what. A recurring statement in the literature is that adding defaults to models affects only actual causation claims and leaves intact the objectivity of causal structure, but there is no obvious reason – apart from assuming realism about causal structure – why counterfactual judgments of actual causation but not counterfactual judgments of causal structure would be affected by what is normal, expected or taken for granted. Unless one recommends definite criteria to mark subjective counterfactuals from those describing objective information, defaults seem relevant to both.[45]

The problem with defaults is twofold: on the one hand, since what counts as a default state is not an objective matter and may vary in a given system depending on circumstances, adding defaults to models affects the scientific, objective notion of causation advanced by the structural equations framework. On the other hand, if default information fails to represent an adequate extension of modelling techniques, the structural equations framework remains with the task of providing a competent definition of what counts as an appropriate causal model. I will briefly discuss each prong.

[44] A cool simulation-based visualisation of how galaxies will collide and eventually merge in a single elliptical galaxy can be found at http://hubblesite.org/newscenter/archive/releases/2012/20/video/a/

[45] As further discussion will make clear, a more accurate statement is that *referential information* – rather than defaults – is relevant to both.

The default-deviant distinction is a notoriously context-sensitive matter. Default states can be natural states of objects or systems, but also states related to norm-dependent contextual parameters. Defaults may also change as a system goes through different circumstances – if a vase is intact, its default state is to remain in one piece, but if broken its default state is to remain broken – and the choice of default states may sometimes seem completely erratic.[46] Therefore, a principle for defining default versus deviant states may prove difficult to pin down. Moreover, despite adapting causal modelling to situations with subtle details, defaults complicate the problem of motivating and justifying actual causation claims, as different modellers may choose different normality orders – statistical, moral, functional – that may lead to inconsistent causal statements. A preliminary step in the justification of default choices would be to enumerate the factors defining normality or typicality – e.g., statistical norms, norms of proper functioning, policies, moral and legal norms – but they cannot guarantee a unique normality theory for a causal model. What we may hope to achieve nonetheless is a framework for the rational evaluation of models (Halpern and Hitchcock 2010); but it is an open question if it is achievable and whether such an achievement would suffice for the scientific notion of causation advanced by Strategy II.

Furthermore, it has been recently argued that default information fails to represent an adequate extension of modelling techniques; instead, attention should be focused on the constraints on what counts as apt modelling with the hope that a careful elaboration of a set of independently justified constraints on variable selection would

[46] Here is an example due to Hall (2007): 'A large rock sits in a sealed room, at noon. Arrayed around the room are sensitive detectors, which will trigger an alarm if they register a sudden pressure change in the room. We ask: what would have happened, at noon, had the rock not been present? That is, what would have happened, had there been no rock in the region of the room where there is in fact a rock? Two contradictory answers are available—each defensible, because each makes tacit use of a different but equally legitimate choice of default state, for that region of the room. First answer: nothing would have happened; so the presence of the rock makes no difference to whether the detectors trigger the alarm. Second answer: without the rock there, a sudden drop in pressure would ensue, as air rushed to fill the empty space; so the presence of the rock is helping to *prevent* the detectors from triggering the alarm. You might find one answer more persuasive than the other. But I think, in fact, that any attempt to rank them is a mistake, which can be brought out by considering this question: What is an appropriate default state for the given region of the room? – A state in which nothing occupies it, one is tempted to answer. That invites a follow-up: Nothing *at all*, or just nothing but what would *normally* occupy it (viz., *air*)? Choose the first answer, and you will judge that without the rock, there would be a sudden drop in pressure; choose the second, and you'll deny this claim. But there is no real conflict here – just a difference between equally acceptable ways of filling in the details of the counterfactual situation that we specify indeterminately as one in which the rock is absent. The example reveals not only a context-sensitivity in the default/deviant distinction, but a way in which that sensitivity can influence *causal* judgments: whether or not we judge the presence of the rock to be preventing the alarm from going off depends on what we take to be the given region's default state'. (Hall 2007: 49-50, emphasis in original).

allow us to arrive at apt models without appealing to defaults (Blanchard and Schaffer forthcoming).

An initial objection is that default-relativity leads to unclarities in cases where the default status is underdetermined or conflictingly overdetermined (Blanchard and Schaffer forthcoming). An example of underdetermined default status is flipping a coin: values for heads or tails cannot be sorted as default versus deviant. In contrast, a default status is conflictingly overdetermined when more than one norm is relevant for defining default/deviant values. For example, if X is driving at 65mph in a 55mph area, what would be the default/deviant values for her speed if the social norm is to drive at 55mph and the statistical norm is to drive at 65mph?

Another concern is that default-relativity does not help in cases of overgeneration, where causal asymmetries (like the gardener's omission/the Queen of England's omission) should rather be treated through independent necessary conditions on model aptness such as Hitchcock's (2001: 287) *seriousness* constraint – 'The variables should not be allotted values that one is not willing to take seriously'. For instance, a model considering the Queen of England's failure to water the plant is described as appealing to a 'too preposterous a scenario' and discarded accordingly. The situation strikes a similarity to Sinnott-Armstrong's (2006) example of an ethical committee dismissing moral nihilism – the possibility that nothing is really right or wrong – as an irrelevant, 'not serious enough' alternative.[47] Following the analogy, causal modellers would need to identify criteria for taking alternatives seriously, or – contemplating Sinnott-Armstrong's line – suspend judgment about the relevance of alternatives. Since meta-scepticism is not an option for the supporters of the seriousness constraint, it follows that unless a justification of this constraint is provided, one is not in a position to declare a model non-apt. Such a justification is not easy to find.[48]

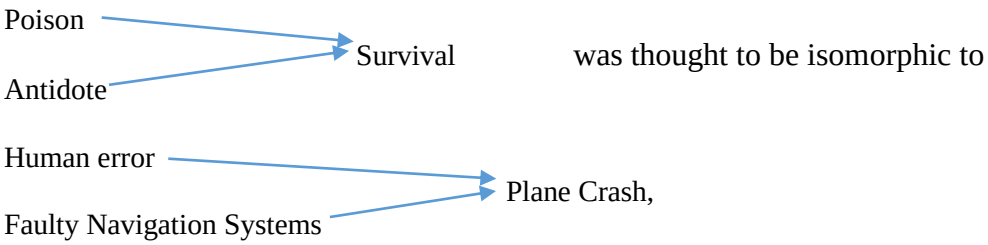
Default-relativity was also supposed to help in cases with isomorphic structures but different causal stories (Hall 2007). In the following two cases,

[47] Similar alternatives are the brain-in-a-vat hypothesis or atheism, e.g. with minds and Christendom apparently having nothing to fear.

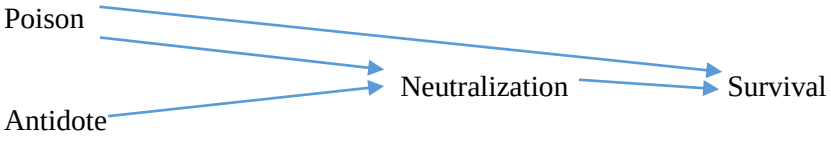
[48] Two remarks:
1) It is important to note that one cannot argue from *evidence* against a relevant alternative like the Queen of England's failure to water the plant. For instance, I might have good evidence to dismiss astrology or theism as irrelevant alternatives because of their inconsistency with other standards of reasoning, but I cannot rule out a causal alternative on evidential grounds.
2) Even if such evidence would have been available, one cannot simply ignore the skeptic. Therefore further argument is needed to show why one could not simply suspend the judgment on whether the gardener or the Queen of England are relevant causal alternatives.

C1: A plane crash (*PC*) is caused by both human error (*HE*) and faulty navigation systems (*NS*).
C2: Killer plans to poison Victim's coffee but eventually refrains (*K*). Bodyguard puts an antidote in the coffee, which would have neutralised the poison (had there been any) (*B*). Victim drinks the coffee and survives (*VS*) (Hiddleston 2005)

the structural equations are isomorphic (HE&NS → PC; K&B → VS), but causally dissimilar: it seems reasonable to understand human error and faulty navigation systems as causes of a plane crash, yet unreasonable to take the presence of the antidote as causing Victim's survival. For that reason, adding defaults such as 'Typically, people do not put poison in coffee' helps the model locate causation in the Killer's act of refraining rather than in the Bodyguard's act. However, these structural isomorphisms are in fact spurious. Once we add a binary variable *Neutralization* describing whether the neutralization of poison occurs, the model is no longer structurally similar to the one in the plane crash example (Blanchard and Schaffer forthcoming). The initial overdetermination model,



but once we add a variable for neutralization, it becomes



In this higher-resolution model we observe that for Antidote to cause the Survival, Neutralization must be {1}, but Neutralization is {0} because Poisoning is {0} (the Killer refrains). Therefore, we arrive at the conclusion that Antidote does not cause the

Survival without appealing to default information.[49] By adding more variables and bringing out the hidden structure of the situation, one shows the initial isomorphism to be based on non-apt modelling. In fact, structural isomorphism is a warning signal for non-aptness and an argument that aptness ultimately falls back on independent constraints such as *Number* – Variables should represent enough events to capture the essential structure of the situation (Hitchcock 2007), or *Stability* – Adding further variables should not capsize causal verdicts (Hitchcock 2010). Yet granting that a move from defaults to constraints on aptness is optimal (at least in certain cases), there is still a long way from the mere 'art' of modelling to a set of objective principles guiding the construction of causal models. Leaving aside the basic necessary conditions – specifying the truth of counterfactuals, an adequate representation of initial conditions, or the inclusion of distinct events – more weighty constraints like *Seriousness*, *Number*, or *Stability* are far from being well-defined and warranting a 'science' of modelling. The structural equations framework needs to provide a competent definition of what counts as an appropriate causal model; it is an interesting open question if the principles guiding the construction of models will turn out to be fully objective.

As a result, contingent on the plausibility of default-relativity, supporters of Strategy II would have to choose between causal modelling with an added context sensitive distinction between the default/deviant states of a system, and causal modelling without defaults but in need of objective principles guiding the choice of variables.

3.6 Can We Justify Modelling Options When Distinct Ways of Representing a System Are Possible?

Even assuming a well-defined set of constraints on model aptness, it is plausible to have multiple apt causal models for a given situation (a determinate system or phenomenon). To begin with, it is possible to have two apt models for a situation – *M1* and *M2* – both using variables *X*, *Y*, *Z*, but one taking $X=x$ to actually cause $Z=z$, and the other taking $Y=y$ to actually cause $Z=z$ (i.e., *M1* and *M2 disagree* on actual causation). If an agent A shoots a Jehovah witness W who once brought to hospital refuses a blood transfusion

[49] More clearly: if Poisoning takes value {1}, Neutralization must be {1}, but it is not; the actual value of Neutralization is {0}. Alternatively, if Poisoning takes value {0}, there is no way to see the causal efficacy of the antidote and therefore its relevance to Survival.

and dies, model *M1* may take A's shooting as the actual cause of death, whereas model *M2* may take W's refusal to actually cause her death.

In addition, it is possible to have two models *M1* and *M2* for a given situation, both arriving at the same conclusion, but appealing to different sets of variables {*A*, *B*, *C*,…} and {*D*, *E*, *F*,…} respectively. For instance, the situation of a child failing at school may be modelled by appealing to variables like {low IQ, culture of poverty, lack of adequate pre-natal diet, economic pressures} or variables like {institutional racism, bad teachers, biased educational standards, lack of appropriate role models}.[50] This possibility raises important questions as to whether *M1* and *M2* compete, reinforce, displace, presuppose, overlap, co-exist at different levels, or are irrelevant to one another. A person who thinks *M1* (or most of its elements) is/are relevant may disagree with the person who thinks that *M2* (or most of its elements) is/are relevant; alternatively, disagreements may crop up with respect to particular variables, e.g., biased teaching vs. low IQ. Consequently, an account of the conflicts or complementarities amongst equally apt models is very much needed.

Furthermore, one may use the same set of variables and arrive at different conclusions. For instance, a causal model of the negative impact of immigration on Britain will allegedly comprise variables such as e.g., {population density, population growth, house prices, unemployment, wages, access to benefits, impact on public services}. Yet the same set of variables is employed to construct a causal model of the positive impact of immigration (e.g., when it is argued that Britain is not overcrowded since only 1.5% counts as developed land, that the number of births and not immigration is the main driver of population growth, that facts do not back up misguided claims about a massive effect of immigration on unemployment, jobs displacement, lower wages, appeal to benefits etc.). The idea is to distinguish between reasoning *from* variables (generating a causal model based on causal evidence) and reasoning *about* variables (assessing the strength and reliability of causal evidence supporting a causal story). Given that it has been recently argued that causal evidence is context sensitive (Suárez 2014), it is an interesting question how a causal model can be manipulated to 'spin' facts or 'push' a story that is influenced politically or otherwise (and thus may alter causal results at will). The point is the following: even if a particular

[50] Adapted from Garfinkel (1981: 2).

modelling option is considered as justified, causal models may include an evaluative dimension, parasitic on or symbiotically attached to them.

Finally, there is a different sense in which two models *M1* and *M2* may compete in the description of a certain system or phenomenon. For instance, a particular segment of reality may be modelled abstracting from more fundamental processes or by describing the processes themselves – e.g., a well-known astronomical event may be represented as the Great Red Spot, or as a persistent anticyclonic vortex in the Jovian atmosphere (that is, as a spot, or as a storm). Similarly, our Solar System can be modelled as being historically stable in human terms, or as being chaotic over billion-year timescales; mountain ranges may be described as particular landforms, or as aspects of a more general process, tectonic geomorphology; and so on.[51] The point is that the variables used in a particular model are defined through abstraction or stabilisation from more basic processes according to criteria in need of justification.[52]

3.7 Models and Responsibility

The discussion so far shows that, obviously, a great deal can be said to answer questions of justification and reduce the arbitrariness in both the construction and selection of apt causal models. However, a lot more work is required to obtain systems of equations accurate and complete in their predictions and interventions. Furthermore, we need a set of principles to deal with the plausible idea that a situation may be modelled in different and sometimes incompatible apt ways, outputting different verdicts about actual causation.

How do these considerations inform moral assessment? Again, as a general point, any pressure on the objective character of causal sequences described by causal models – or any doubts about the structural equations framework as an adequate and non-arbitrary way of capturing the dependence relations amongst phenomena – will

[51] A somewhat similar worry is expressed by Ned Hall: 'In addition, it may not always be straightforward how to "carve up" the given system into sub-systems. It will be *fairly* straightforward, if the system is constituted by a number of clearly distinguishable, interacting parts. But that won't always be the case—at least, at the desired level of description. Consider the flow of water down some rapids: what choice could we make of interacting parts, given that we don't wish to introduce variables for the state of each water molecule at each moment? Here a kind of default option suggests itself, which is that we choose variables to correspond to reasonably well-defined regions of space at different times, or regions of spacetime. The price of exercising this option is, in general, that no set of variables will stand out as uniquely appropriate.' (Hall 2007: 13, emphasis in original).

[52] Again, it is an interesting methodological question whether causes may be thought of as, and substituted with, causal paths (cf. Swanson [2010] for a similar suggestion).

affect the views attempting to tie attributions of responsibility to an objective relation. In particular, if outcome responsibility attributions are grounded by causal sequences represented in causal models, each of the problems associated with the justification of modelling choices will turn out to be relevant to moral assessment. For example:

a) If adding more variables overturns causal verdicts in a model, responsibility attributions may follow this variation (as shown in Section 3.2). One will therefore need a constraint such as *Stability* to conserve both causal and moral verdicts. But if *Stability* seems to be a mere rule of thumb when it comes to determining the right number of variables to be included in a model, there are weak hopes that one could offer 'a more complete mathematical characterisation of the conditions under which the verdicts of actual causality remain stable under the addition of further variables' (Halpern and Hitchcock 2010: 395). Absent a clear stability requirement, a model may misrepresent the causal evidence on which responsibility attributions are based.

b) I mentioned that supplementing the formalism of causal models with default information (information about what it is normal, expected or taken for granted with respect to the states or evolution of a system) will help determine what (actually) caused what, as we would judge something to be a cause if it deviates for the presumed default state of the system. In Collingwood's example (Section 2.1), the situation may be described by appealing to four variables:

A for the accident, where $A = \{1\}$ if there is an accident and $A = \{0\}$ otherwise.
CF for cornering too fast, where $CF = \{1\}$ if the driver cornered too fast and $CF = \{0\}$ otherwise.
DS for the defective surface of the road, where $DS = \{1\}$ if the road was defective and $DS = \{0\}$ otherwise.
DD for the defective design of the car, where $DD = \{1\}$ if the design was defective and $DD = \{0\}$ otherwise.

Adding default values, we expect or take for granted that accidents should not happen, so the default state on the effect side is 'no accidents'. On the cause side, we expect drivers to corner slowly, roads to be impeccable, and cars to be soundly designed. However, as Collingwood points out, if the county surveyor takes the default state of

the system to be 'flawless roads', she will see the actual cause of the accident to be *DS* = {1}; if the motor manufacturer takes the default state of the system to be 'the sound design of the car', she will see the actual cause of the accident to be *DD* = {1}; and so on. Clearly, a variation in what one takes the default information to be will affect actual causation verdicts and therefore the responsibility evaluations based on them. If the default deviant distinction is a highly context sensitive matter (*as per* Section 3.4), the context sensitivity may reflect on moral assessment (for a detailed discussion of this possibility, see Chapter II).

c) Perhaps the most dramatic variation in moral assessment comes from the possibility of having multiple appropriate causal models for a given situation. Two similarly apt models may use the same variables in representing the causal structure of a situation, but disagree on actual causation, and therefore on moral responsibility (recall the Jehovah witness case in Section 3.6: a model *M1* may settle on the shooting as the actual cause of death, while model *M2* may settle on the blood transfusion refusal as the actual cause of death). Moreover, two similarly apt models may use different variables and arrive at different representations of causal facts and therefore at different moral verdicts. For instance, causal evidence may be differently interpreted in a court of law and used to support distinct moral verdicts. Furthermore, two similarly apt models may use the same set of variables, but interpret them differently such that causal facts may be 'spun' and a particular responsibility story 'pushed'. For instance, certain data may be exaggerated or presented with rhetorical force to manipulate and achieve strong demagogic effects. Lastly, if causal facts are modelled at different levels of abstraction, responsibility claims may be associated with each theoretical level or reduced to the most fundamental one – for instance, both groups and their members may be said to be responsible for a certain outcome in the world; alternatively, a group's responsibility may reduce to the responsibility of its members. But attention to the theoretical level a model is pitched seems to be very important, especially in cases of scapegoating (where the responsibility of many is left with a few), responsibility dissipation (where the responsibility of few is left with the many), or 'suspended' responsibility (where the genuine locus of responsibility for an outcome is uncertain and left to bounce between theoretical levels). I will return to the discussion of higher-level vs. lower-level causation and causal overdetermination in Chapter III.

3.8 So Far

Recall that our intuitive conception of causation relies on incompatible intuitions competing for priority and one way of addressing this incompatibility was through a *reconciliatory project* which defends the objectivity of causation and explains perspectivist intuitions as being explained by the shiftiness of our causal talk. A different way of approaching our incompatible intuitions, discussed in the second part of the chapter, favoured a *utility-based project*: one may *choose to define* a scientific concept of causation used for prediction, intervention or rational deliberation. The fact that such a concept is theoretically useful may well compensate for the lack of fit with certain intuitions. Strategy II acknowledges the presence of incompatible intuitions, but redirects those of a realist persuasion to focus on the causal modelling framework as the best theory out there delivering a scientific, hence non-arbitrary (or less arbitrary) conception of causation. However, such a conception faces allegations of arbitrariness at several levels – most importantly related to the choice of variables, their values, the notion of default information vs. objective constraints on aptness, and the selection amongst competing but equally apt causal models – and I have suggested that the criteria involved in the construction and selection of apt causal models will ultimately be pretty thin. If I am right, the considerations against model justification will reflect on the scientific notion of causation argued for, and by extension, on the views which hope to tie attributions of outcome responsibility to such an objective relation.

4. Conclusions

Most theorists interested in how causation and responsibility relate take causation to be an objective and mind-independent relation between space-time located relata. A key advantage of such realist sympathies is that an objective, extensional causal relation would provide moral assessment with a naturalistic basis, making moral properties continuous with a scientific view of the world: it is in virtue of a determinate feature of reality that we take responsibility claims to be genuine. But the uncertainties surrounding the project of defending the objective nature of causation challenge the realist tendencies. They show that both describing and modelling of causal facts are affected by ambiguities, and as a result, realists need to retreat towards more modest or measured commitments. Therefore, an analysis of one important feature of causal

relations – their objective reality – has led us to question views of responsibility aiming to provide moral assessment with a naturalistic underpinning. The analysis also showed that we are in fact confronted with a difficult issue: we tend to think of causal relations realistically, but most of our intelligible causal claims do not easily fit this mould. Causation *in fact* may often prove to be causation *to us*.

The next chapter takes up a different aspect of causal relations, i.e. their context sensitivity. It explores the connection between the context sensitivity of causal claims and the intuitively plausible idea that moral responsibility rests on causation, providing further insight into the way certain complex features of causation may impact our current perspective about responsibility claims.

II. Moral Lessons from the Context Sensitivity of Causal Claims

'…[I]n both law and morals the various forms of causal connection between act or omission and harm are the most obvious and least disputable reasons for holding anyone responsible.' (Hart and Honoré 1985: 66)

Introduction

Context influences causal claims. For instance, examining what caused the Tyrolean Alps avalanches of December 1916, the WWI historian might cite the artillery fire, while the glaciologist might cite slope angle and orientation, shockwaves, or snowpack structure. Noting this variation, Lewis surmises:

> We sometimes single out one among all the causes of some event and call it "the" cause, as if there were no others. Or we single out a few as the "causes," calling the rest mere "causal factors" or "causal conditions." Or we speak of the "decisive" or "real" or "principal" cause. We may select the abnormal or extraordinary causes, or those under human control, or those we deem good or bad, or just those we want to talk about. I have nothing to say about these principles of invidious discrimination. (Lewis 1973: 558-559).

The context sensitivity of causal claims has received sustained attention in the literature (Mackie 1980, Lewis 1973, Bennett 1995, Hitchcock 1996, Woodward 2003, Maslen 2004, Menzies 2004, Swanson 2010, Schaffer 2012), but it has not been discussed in connection to an intuitively plausible idea, *viz*. that moral responsibility claims rest on causal claims. If responsibility attributions rest on the assessment of causal sequences relating agents, events, and consequences, it is legitimate to ask if moral evaluation inherits the context sensitivity of causal claims. I argue that taking context sensitivity seriously generates important worries about moral responsibility.

Overview: I start by defending the idea that moral responsibility rests on causal sequences (Section 1). After presenting causal contextualism (Section 2) and examining the ways it impacts moral assessment in complex situations (Section 3), I take a look at some alternative devices for stabilising causal thinking in moral contexts and show why their applicability is in principle restricted (Section 4). I contend that we end up with a choice between moral contrastivism and a mild version of scepticism, *viz.* moral responsibility is not impossible, but ultimately difficult to identify with confidence. I show why the latter view is more plausible (Sections 5-6). What emerges is that the more we know about the causal complexity preceding a particular event, the more uncertain we become about what actually causes it. As a result, tying responsibility assessment to causation may bring uncertainty to moral evaluation.

1. Causation and Responsibility

1.1 The Standard View

We are responsible for our acts and their outcomes (events and states of affairs in the world).[53] Since the only way to make a difference in the world is through our causal powers, the moral assessment of our acts is intuitively related to what we cause. For instance, I am not responsible for overfishing green sea turtles in the Pacific as there is no causal connection between me and the action of pirate trawlers. This argument from connection is usually accompanied by an argument from moral phenomenology: we experience stronger moral feelings when a person engages in culpable behaviour causing harm (e.g., an intoxicated driver skidding and hitting a child) than when a person engages in a similar culpable behaviour but does not cause such harm (e.g., an intoxicated driver colliding with a tree).[54] The standard view (**SV**) follows naturally:

SV If an agent A is morally responsible for an outcome O, A must have performed an action that caused O.[55]

Two clarifying remarks are in order.

First, the standard view may take more versions. For instance:

[SVversion1] If an agent A is morally responsible for an outcome O, A must have performed an action that was *the cause* of O.

[SVversion2] If an agent A is morally responsible for an outcome O, A must have performed an action that was *a cause* of O.

[SVversion3] If an agent A is morally responsible for an outcome O, A must have performed an action that *causally explains* O.

[53] The focus here is on the latter. Of course, outcomes refer to an agent's acts and an agent's acts require a discussion of notions such as freedom or control, however, these issues will not be addressed here. I will also leave aside the views attaching responsibility to beliefs rather than choices, actions or outcomes (e.g. Moya 2006).

[54] Although see Gardner (2012) for some criticism of this so-called experiential argument.

[55] **SV** also appears in the literature as the' entailment claim' (Sartorio 2007, Driver 2008). The stipulation surfaces in Heinaman's (1986) account, doubled by an explanatory claim. As Fischer and Ravizza remark, '[o]n Heinaman's approach, an agent is morally responsible for a state of affairs insofar as he knowingly and intentionally does something, the doing of which at least in part explains why the state of affairs obtains.' (Fischer and Ravizza 1998: 104). A more refined version of the **SV** includes the idea that causation grounds responsibility, that is, it determines, explains and makes true responsibility claims.

Considering [SVversion1], 'A *caused* O' is most of the times interpreted as 'A is *the cause* of O': as Lewis (1973) noted, we have the tendency to single out or elevate certain factors to the status of cause, while relegating other factors to the status of background conditions. In general, responsibility attaches to agents as causes in the sense conveyed by [SVversion1].

Considering [SVversion2], an agent A is '*a cause* of O' is used either when multiple agents cause an outcome, when supplementary causal factors, beside the agent, are relevant to an outcome, or simply when a multitude of different factors cause an outcome. On the one hand, being *a cause* seems a less context sensitive notion, but even to elevate something to the status of *a* relevant cause – from amongst *all* potential causes – still takes some selection and is thus still context sensitive, albeit in a different sense. On the other hand, being *a cause* is less informative. We do not usually stop at remarking the causal complexity of the world: we want to know *the cause*. There seems to be an interesting relation of indirect proportionality between context sensitivity and informative character, however, it cannot be taken up here for reasons of focus and space.

Considering [SVversion3], 'A *caused* O' could be replaced with 'A must have performed an action that *causally explains* O'. Such a move is helpful – for instance, it would allow omissions to causally explain certain outcomes without granting them genuine causal power. However, the success of this move depends on how the debate on negative causation is resolved. If it surfaces that omissions have causal power, there is no need to replace causation with causal explanation. In addition, replacing causation with causal explanation will not fend off context sensitivity, as variation is retained at the level of causal explanations.

In light of these considerations, there are two options regarding the versions of **SV**:

a) Consider 'cause' an umbrella term for *the cause*, *a cause* and *causally explains* and use the initial formulation of the standard view.

b) Revise the initial standard view to include omissions and end up with a disjunctive formulation:

SV [revised]: 'If an agent A is morally responsible for an outcome O, A must have performed an action or omission that caused *or* causally explains O'.

However, since each of the suggested interpretations – *causes, is the cause, is a cause* and *causally explains* – is plausibly a context sensitive notion,[56] I remain neutral between **SV** and **SV [revised]** and focus the discussion on the context sensitivity of 'caused' as it appears in **SV**. In different terms, **SV** may take different interpretations, but since all causal notions involved are context sensitive, the argument will be centred on the initial formulation: if an agent A is morally responsible for an outcome O, A must have performed an action that *caused* O.

Second, some will be quick to note that responsibility ascriptions also depend (or fundamentally depend) on agency-related conditions such as intention, control, foreseeability, epistemic competence, or psychological consistency. To be morally responsible in this sense is to be a sound agent, i.e., to be perfectly aware of the options presented by a particular situation, to act according to principles, to control one's actions and understand their results, to foresee their consequences, etc.

However, if attention is restricted to responsibility for outcomes in the world rather than for actions or beliefs, the causation requirement takes analytical priority over the sound-agency condition: the former must be satisfied before the latter. With few exceptions – e.g., strict liability cases – responsibility is ascribed to persons as causes of events or states of affairs, with further questions related to intentions, reasons, or control arising after a relevant causal connection is established. Of course, we often ascribe responsibility for actions or mere intentions. But responsibility for actions is mitigated or annulled when no relevant consequences appear (assuming that responsibility for inchoate offenses is not on a par with outcome responsibility). Similarly, responsibility for intentions is often meaningless if simple intentions are not followed by action and consequences.[57] In addition, it might be that other agency-related conditions could be re-described in causal terms. For instance, the ability-to-do-otherwise condition (Frankfurt 1969) – the idea that a person is morally responsible only if she could have acted otherwise – could be re-described in terms of alternate opportunities to break the existing causal flow of events, or in terms of alternate chances to make a difference in the world.

[56] Although some are more context sensitive than others.
[57] For instance, cases of coercion are usually invoked to dissociate responsibility from caused outcomes. But it is often overlooked that it is outcomes that prompt assessment questions about the justified or unjustified behaviour of the coerced person, about fault vs. exoneration, and so on.

The causation requirement is also important because sometimes one is morally responsible for an outcome irrespective of whether that outcome was unintended, uncontrolled, or unforeseen. Lastly, the causation requirement matters because outcome allocation is essential to our identity as persons. For instance, Honoré remarked that 'having decided nothing and done nothing [we] would hardly be people' (Honoré 1988: 543), a point which was further developed by Gardner:

> [A]nyone who denies the existence of moral reasons to bring about results or consequences also denies the existence of (normal) moral reasons to *try to* bring about results or consequences. So if one will not assess actions morally according to their results or consequences, one should not assess them according to their intentions either. If our successes do not count, then neither do our endeavours (Gardner 2012: 461, emphasis in original).

Therefore, there are reasons to insist on the relevance of causation to moral responsibility assessment (as least as far as outcomes are concerned). However, to do justice to all conditions (intention, control, foreseeability etc.) and at the same time maintain the focus on causation, I will assume a composite conception of moral responsibility, *viz*. one that takes causation to be one of the core elements of responsibility, next to intention, control, adequate reasons, psychological consistency, or normative pressure towards an answer (for instance, the answer demanded by the legal norms). Since all these elements are important, moral responsibility is a composite notion.[58] However, the emphasis placed on its elements always varies: there is no such thing as an appropriate way of adjusting them, one 'correct' conception of responsibility, but different interpretations based on how we balance its core elements. When outcome responsibility is under assessment, causation becomes the axis of moral evaluation rather than being demoted to the status of a mere secondary element, one amongst the many necessary conditions of moral responsibility.

With these remarks added, let me return to the standard view – if an agent A is morally responsible for an outcome O, A must have performed an action that caused O. Despite its intuitive appeal, **SV** was considered a somewhat limited position. According to its critics (Beebee 2004b, Dowe 2004, Sartorio 2004, 2007 *inter alia*), various counterexamples could be called upon to show that being responsible for an outcome does not necessarily require causing it. Such counterexamples include cases involving omissions, situations of causal overdetermination, or circumstances in which moral responsibility explains causation (and thus reverse the explanatory order transmitted by

[58] A similar view is suggested by Williams (1993).

SV). Next, I shall argue that such counterexample-based objections are not strong enough to motivate a revision of **SV**.

Objection 1: Omissions

An initial objection starts by pointing out that we are often responsible for omissions. For instance, my omission to switch off my mobile phone causes the disturbance of a fine Schubert evening. However, since omissions are literally nothing and 'nothing from nothing ever yet was born',[59] they cannot be causes. If we are morally responsible for omissions, and omissions are not causes, then one can be morally responsible without being causally responsible. To some theorists, ordinary discourse involving negative causation is rather guided by the aims of explanation (Beebee 2004b), or it is at best quasi-causal (Dowe 2004), but there is no such thing as negative causation *per se*.[60]

The idea that omissions are 'nothings', non-actions lacking genuine causal efficacy or 'oomph', has considerable intuitive support. But this should not dismay **SV** supporters, as the standard view could be revised to include omissions. For instance:

[Omission-friendly SV] If an agent A is morally responsible for an outcome O, A must have performed an act (action or omission) that caused O.

In addition, there are reasons to treat omissions on a par with causes. For instance, denying omissions causal efficacy would entail a radical abandonment of common sense, as negative causation appears in most of the paradigm cases of causation considered by our ordinary, moral, and legal language, including all human actions,

[59] Titus Lucretius, *De Rerum Natura*, Book 1. Translated by William Ellery Leonard, Internet Classics Archive, available at http://classics.mit.edu/Carus/nature_things.1.i.html

[60] To Beebee (2004b), causal claims involving omissions or absences as relata are not really causal, but merely explanatory. To Dowe (2004), omissions are best explained in counterfactual terms, for instance the thought that my failure to help a person in need *caused* her death could be rendered as 'had I helped that person, she would not have died'. Since counterfactual claims are quasi-causal statements referring to possible actions, omissions are not causal but rather quasi-causal. Obviously, Dowe is not a supporter of a counterfactual theory of causation. To him, real causation connecting concrete occurrences needs to be distinguished from quasi-causation (expressed with the help of counterfactuals). The metaphysical status of omissions cannot be taken up here in detail for reasons of space, but it must be noted that doubts concerning the causal power of absences, omissions, or other negative events force some theorists to adopt a *disjunctivist* account of the grounds of responsibility (e.g., Moore 2009). These authors hold that moral responsibility is grounded *either* in causation *or* in relations of mere counterfactual dependence. In general, siding against negative causation are authors such as Alvarez (2005), Beebee (2004b), Dowe (2000, 2004), Lewis (2004), Moore (2009), Thomson (2003), or Varzi (2007). Arguing for the causal efficacy of negative events are authors such as McGrath (2005), Hunt (2005), or Schaffer (2000, 2004).

biological functions, or mechanical movements (Schaffer 2004). To illustrate, if X is releasing the lever of a catapult and sends a brick through the window, there is *nothing* connecting the release of the lever and the flight of the brick – 'no relevant energy-momentum flow, track of mark transmission, or persisting trope' (Schaffer 2007). However, no one would deny that X caused the window to break or that X is responsible for it. For that reason, it looks like **SV** would need to include omissions in any of its plausible versions.

The main worry for negative causation supporters is causal overgeneration. Suppose that I forget to close the French doors leading to my rear garden and I find a fox in my lounge. Intuitively, my failure to close the doors caused the presence of the fox in my lounge. If my failure is causally efficacious in this sense, and a non-action remains a non-action no matter to whom it belongs, then *any* non-action of *any* remote agent X, Y, Z is equally efficacious and equally relevant. For example, Morgan Freeman's failure to close the doors is an equally good causal answer to explain the presence of a comfortably coiled up fox in my lounge. The question is then how can we metaphysically distinguish the relevant omissions from the irrelevant ones? There are in principle various strategies to avoid causal overgeneration. To single out the relevant omissions one may appeal to pragmatic considerations (Bennett 1995, Lewis 2000), contrast values (Schaffer 2005), or independent normative standards (McGrath 2005). Of course, these strategies may end up being inconclusive. But even if all arguments for granting omissions genuine causal efficacy or for treating them on a par with causes fail, the discussion will move to the context sensitivity of causal explanations. More precisely, if omissions cannot figure in causal claims but can safely figure in causal explanations, accommodating omissions would lead to the replacement of **SV** with [SVversion3] – if an agent A is morally responsible for an outcome O, A must have performed an action that *causally explains* O. But at this point critics of negative causation such as Beebee or Dowe would have to argue that causal explanation itself is not grounded in an objective relation, that is, they would have to show that an event C does not explain an event E in virtue of an objective, determinate relation in the world, but in virtue of a logical, conceptual or epistemic relationship between items of knowledge. Without going into details, it is safe to say that in order to avoid anti-realist commitments, most theorists of explanation would argue that causal explanations are underpinned by genuine causal sequences 'out there', which takes us back to **SV**.

Objection 2: Transmission

Another objection to **SV** invokes cases of preemptive overdetermination.[61] Consider McLaughlin's example: 'Suppose A is entering a desert. B secretly puts a fatal dose of poison in A's water keg. A takes the keg into the desert where C steals it, thinking that it contains pure water. A dies of thirst. Who killed him?' (McLaughlin 1925: 25). This seems a case of responsibility without causation – it is clear that someone is responsible for A's death – B, C, or both. But none of them seems to be a cause, as causal chains cancel each other out. However, it could be argued that since each individual is responsible for their joint behaviour, which in turn caused the outcome, each of them is responsible for the outcome. In this sense, causation is *transmissible*, or functions as a vehicle of responsibility: being responsible for a cause of a particular result makes one responsible for that result (Sartorio 2007). Such examples aim to motivate a slight revision of **SV** rather than fully object to its convincing character. To be morally responsible one does not necessarily need to perform an action or omission causing an outcome O; it is enough to be part of a joint behaviour which in turn causes that outcome. In this sense, moral responsibility will not automatically *entail* causal responsibility (as **SV** holds), but causation would still *ground* moral responsibility attributions. Cases of preemptive overdetermination seem to make apparent a limitation of **SV**. But the objection is not decisive.

First, note that transmission cases are still based on causation, but the problem is made a bit more complicated by introducing the idea of joint action and distributed responsibility. In different terms, moral responsibility still requires an act (action or omission) causing the outcome, but now the act is joint or collective. Moreover, both the entailment and the grounding claims could be subsumed to the more general idea that moral responsibility *rests* on causation (although this might also be read as a metaphor for grounding).

Second, transmission contravenes the plausible principle of causal-moral proportionality: intuitively, an agent's moral blameworthiness for an outcome is proportional to the agent's causal responsibility for that outcome. A Wall Street broker who sold high risk, complex financial products and a statistics analyst in a credit rating

[61] In cases of preemptive overdetermination, causes are temporally ordered. This type of overdetermination is different from concurrent cause overdetermination (simultaneous causes) or asymmetrical overdetermination (simultaneous causes, but having a distinct weight and contributing asymmetrically to generating the outcome).

agency are both morally responsible for their causal contribution to the Global Financial Crisis, but not for the entire crisis. However, if the principle of transmission is correct, being part of a joint behaviour would make a person disproportionately responsible.

Third, leaving aside the restricted reliance on joint action, there are counterexamples to the claim that agent A is responsible for an outcome O iff A is responsible for a cause of O – e.g., the Doctrinal Paradox identified in the jurisprudence literature by Kornhauser and Sager (1993).[62] Although this type of dilemma is used to emphasize the difficulties in aggregating individual opinions in a coherent collective view, one could use such paradoxes to show that a person can be responsible for a cause of an outcome *without* being responsible for that outcome (and thus block the idea of transmission). Consider a three-member court evaluating a case of liability for a breach of contract, where liability depends on assessing the defendant's action and her contractual obligation not to act. The judges have to decide on the truth of the following propositions:

P: The defendant did a particular action.
Q: The defendant had a contractual obligation not to act as she did.
R: The defendant is liable.

All judges agree that $R \equiv P \wedge Q$. Next, they can either decide the issue by appealing to majority voting on each premise P and Q (*the premise-focused procedure*, or aggregating votes with respect to the conditions of liability), or by voting directly on the conclusion R (*the conclusion-focused procedure*, or aggregating votes with respect to liability itself). However, as appears in the Table 1 below, the resulting judgments are inconsistent:

	P (Action?)	Q (Obligation?)	R, P∧Q (Liability?)
A	Yes	Yes	Yes
B	No	Yes	No
C	Yes	No	No
Majority Decision	Yes	Yes	**Yes/No**

Table 1. The Doctrinal Paradox

[62] The Doctrinal Paradox (Kornhauser and Sager 1993) refers to cases of collective decision-making through majority voting in which groups must decide on a complex issue depending on other previous issues. The paradox highlights the conflict between the majority vote on the complex issue itself and the majority vote on the other previous issues the complex issue depends on. It aims to show that responsibility for a cause of an outcome can be completely independent from the responsibility for the outcome itself – which here undermines Sartorio's idea of transmissibility.

If the judges aggregate votes with respect to the liability (R), and liability (R) depends on both action (P) and contractual obligation (Q), it appears that two out of three judges believe that the defendant is *not liable* (B, C). However, the defendant would be found *liable* if judges aggregated votes on the previous issues, action (P) and contractual obligation (Q), because there is a majority on each previous issue, P (A, C) and Q (A, B). The Doctrinal Paradox emphasizes that procedures lead to incompatible results. Take the conclusion-driven procedure and the conclusion will be inconsistent with the majority on previous issues. Take the premise-driven procedure and the judges will support a conclusion that a majority of them individually rejects. Either way, responsibility for the individual decisions seems *discontinuous* with the responsibility for the outcome (here, the majority decision), even if these decisions *caused* the outcome.[63]

Such situations are widespread and they appear anytime a group of people reason together on multiple premises with the aim of forming an opinion on a complex issue. If these counterexamples work, they block the idea of transmission introduced by preemptive overdetermination cases: one can be responsible for a cause of an outcome without being responsible for that outcome.[64] Consequently, the argument from transmission is not sufficient to motivate a revision of **SV**.

Objection 3: Grounding vs. Explanation

The third objection targets **SV** indirectly. Granting that **SV** is plausible and causal sequences ground moral responsibility attributions, one could still observe that sometimes moral responsibility *explains* causal responsibility according to the principle

[63] To make the case stronger, one could adjust the voluntary or epistemic constraints of responsibility – suggesting, for instance, that the judges are not morally responsible for the outcome (but only for a cause of it) because they are rather following a certain procedure and do not know what the outcome will turn out to be. In reply, one might claim that judges are still morally responsible for the outcome even if some of them disagree with it, since they have responsibilities associated with their office or position. However, there is a difference between their moral responsibility for a particular decision and their moral responsibility associated with their roles. The idea is that when ascriptions of responsibility presuppose neglecting intention or causation, this happens because there is an assumption of responsibility associated with the role, rather than to a certain decision. If one focuses on the role, outcome responsibility would in principle attach to all decisions and – *in extremis* – would preclude the possibility of dissent. But if we focus on a particular decision, it is possible – at least in principle – to hold that one is responsible for a cause of the outcome, but not responsible for the outcome itself (against Sartorio).

[64] The Doctrinal Paradox has been used to argue that collective agents can be morally responsible for an action or decision without their members bearing such responsibility (Copp 2006, Pettit 2007 *inter alia*). I am not persuaded by these arguments and it is not my intention to claim that cases of causal overdetermination or joint determination entail a notion of moral responsibility independent from individual responsibility; my sole interest lies with pointing out the limitations of the transmission view.

'If x is at fault for y, then x caused y' (Thomson 2003). So perhaps **SV** is only partially correct and should be revised to include cases when we ascribe causation because we have previously ascribed responsibility and not the other way round. Including such cases would shed light on our psychological tendency to hold agents causally responsible because we seem already convinced of their moral responsibility. In addition, the observation concerning the direction of explanation would help us distinguish causal from non-causal omissions – e.g., since it was *my* responsibility to help the person in need, *my* omission caused her death (and not the omissions of my 'brothers in inaction' X, Y, Z).

However, assuming that **SV** is plausible, the direction of explanation should mirror the direction of grounding – that is, if causal sequences explain responsibility ascriptions, this happens because causal connections also ground our responsibility ascriptions. By reversing the explanatory order, the last remark indirectly casts doubt on the grounding relation: what if the presumed objective analysis of the causal relations underpinning moral responsibility attributions contained in fact a hidden normative parameter, intimately related to human purposes and concerns? The answer is that we do not know whether causal analysis will ultimately turn on normative considerations (although e.g., Mackie [1955], Alvarez [2005], Thomson [2003], and McGrath [2003] point sometimes in that direction). But given that explanatory practices run both ways (causation explains responsibility and responsibility explains causation), one could offer at least two answers of principle to this third objection:

(**A1**) In principle, even if sometimes causation seems dependent on and directly proportional to the strictness of normative requirements (especially in the legal domain), we do not normally derive causal conclusions from moral premises. In principle, normative notions should not be employed to settle matters in the natural domain (where causation belongs).[65]

[65] There is no consensus on the metaphysical status of causation: some argue that causation is part of the fundamental fabric of reality, some think it is relatively non-fundamental and analysable in terms of other basic traits (e.g., regularities, laws, energy), and still others think that it is a merely projected feature on what is happening in the world. Without engaging in a detailed defence of a side, it is worth noting that even if sometimes causation seems projected onto the world because of its connection to human purposes and concerns, other times it is not – for instance, when it applies to the migration of Canada geese or to the weather cycles in the Mont Blanc massif.

(**A2**) In principle, one should start with a separate account of causal responsibility and then assess moral responsibility in terms of the initially considered causal sequences. Pondering whether the guest or the cat is responsible for breaking the vase, we evaluate their causal responsibility *in the same way*, i.e. by determining the sequences of events connecting their acts to the outcome. We would not explain the causal responsibility of our guest in terms of her moral responsibility simply because she can be morally responsible and the cat cannot. We would not treat them differently (Fischer and Ravizza 1998, Sartorio 2007).

In this section, I introduced, motivated and defended the standard view of the relation between causation and responsibility (**SV**) – if an agent A is morally responsible for an outcome O, A must have performed an act (action or omission) that caused O. If correct, **SV** emphasizes the deep-seated relation between the structural features of action and the practice of attributing praise and blame. If causation grounds and explains responsibility attributions, the way we understand causal claims will certainly impact moral evaluation.[66] Next, I offer some examples of the context sensitivity of causal claims[67] and briefly discuss its sources (Section 2). I then examine its implications for responsibility attributions (Section 3).

2. The Context Sensitivity of Causal Claims

2.1 Definition and Illustrations

Causal claims are context sensitive in the sense that the truth or acceptability of causal statements of the form 'C caused E' (or employing similar causal verbs) varies across contexts. For instance, considering what caused the 2008 K2 climbing disaster, A might mention human error, B the collapse of a massive serac in the notorious Bottleneck couloir, C the high altitude sickness, and D the faulty gear – and what is accepted as a cause by a speaker may be genuinely denied by others. Although the context sensitivity

[66] Recent empirical research in moral psychology distinguishes two processes of moral evaluation, one based on the assessment of causation and the other based on the assessment of intentional factors. Judgments of responsibility, blame and punishment were found to rely on causal connections between agents and outcomes, while judgments of wrongness and permissibility of action were found to be associated with the mental states of the agent. Developmentally, the causation/consequence-based psychological system for moral judgment comes first, and it is later augmented by a different mental-states-based system (Cushman 2008).

[67] The context dependence of causal claims is discussed by Mill (1843/1947), Mackie (1980), Lewis (1973, 1986b), Hitchcock (1996), Woodward (2003), Maslen (2004), Menzies (2004), Schaffer (2005, 2012), or Hall (2007) in the philosophical literature, but it was also approached in early sceptical accounts of legal causation (e.g., Malone 1956, Borgo 1979, Landes and Posner 1983).

of causal claims is a largely uncontroversial idea, it received little attention in the literature if compared to the context sensitivity of knowledge attributions. The most discussed examples of context sensitivity are associated with the context of occurrence and the context of inquiry:[68]

(i) Consider the selection of causes from background conditions. If a house is destroyed by fire, one cites the short circuit as the cause of the fire and takes the presence of oxygen as a mere background condition; but if a fire breaks out in a lab where precautions of excluding oxygen are taken, the presence of oxygen is cited as the cause. The presence of oxygen counts as a cause or condition depending on the context in which the fire occurs. In general, different situations generate different claims about what count as causes and background conditions.

(ii) Context sensitivity is also associated with "who asks the question and why" (Hart and Honoré 1985). If I shoot a wild boar with my .270 Winchester and the question is what caused me to shoot it, the answer must specify my interest in shooting wild boars rather than, say, red deer. If the question is why I used a .270 Winchester, the answer must stipulate why I have chosen this rifle and calibre rather than, say, a Remington 700 CDL chambered in .35 Whelen. Finally, if the question is why I shot the wild boar rather than hugging it, the causal answer must identify my attitude towards wild boars or the circumstances rendering boar-hugging inconceivable. It follows that causal claims – which in this example are built in the description of action – are sensitive to the interests governing causal inquiry.

Beside these most often mentioned instances of context sensitivity, the idea is further illustrated by an ample spectrum of cases. There are, for instance, cases involving deviant causal chains,[69] preemptive preventions,[70] cases displaying

[68] The terms belong to Gorovitz (1965: 700).

[69] Consider Davidson's example: 'A man may try to kill someone by shooting at him, (…) misses his victim by a mile, but the shot stampedes a herd of wild pigs that trample the intended victim to death.' (Davidson 1980c: 78). Causal judgments tend to vary because different contexts bring out distinct features of the situation as relevant for the outcome, i.e. the shooting, the missing, the presence of wild boars. If one may be inclined to claim that the man's shooting caused the victim's death, others may think that the wild boars caused it – after all, the man's shooting missed the victim by a mile.

[70] Consider McDermott's example: 'A cricket ball is hit with substantial force towards a window. A fielder reaches out and catches the ball. The next thing along in the ball's direction of motion is a solid brick wall. Beyond that is the window. Did the fielder's catch prevent the ball hitting the window?' (McDermott 1995: 525). Depending on context, one has the tendency to think both that the fielder's catch does not make any difference to the hitting of the window *and* that his catching does prevent the ball from hitting the window. As a result, causal judgments vary depending on the intuition we are inclined to trust.

alternatives and contrasts,[71] description shifts and focus shifts,[72] or other circumstances (for additional examples, see Menzies [2007], Schaffer [2012]).

2.2 Sources: Semantics vs. Pragmatics

In general, it is debatable whether context sensitivity is a question of semantics or conversational pragmatics – whether it is encoded in the meaning of causal statements or reflects the interests and expectations specific to the conversational point.[73] The literature is divided on the issue. A pragmatic view is favoured by Mackie (1980), Lewis (1986a, 1986b), Bennett (1995), and Swanson (2010). A semantic view is

---

[71] For instance, suppose that I could reach Calais from Dover by car, ferry, train, or plane. I choose to travel by ferry (although unbeknownst to me, the Eurotunnel is temporarily closed). Did my choice cause my arrival in Calais? The answer is *yes* if the expectation in play or conversational assumption is that I could have travelled by train (as I often do), but *no* if the expectation is that I could have flown there: in the latter case, flying or taking the ferry makes no difference as one expects me to arrive at my destination anyway. This type of context sensitivity can be also rendered through manifest contrasts stipulating the options, such as 'My taking the ferry *rather than* travelling by train caused my arrival in Calais *rather than* my being stranded in Dover.'

[72] The idea is that shifting description or focus changes the acceptability of causal claims. Here are two examples for description- and focus-shift respectively (Schaffer 2012):

1. McEnroe's tension caused him to SERVE.
2. McEnroe's tension caused him to SERVE AWKWARDLY.

A person with a fine-grained conception of event individuation denies 1 and accepts 2, whereas a person with a (more) coarse-grained conception of event individuation sees one action and holds that serving just is serving awkwardly. Further,

3. Socrates's DRINKING HEMLOCK at dusk caused his death.
4. Socrates's drinking hemlock AT DUSK caused his death.

Again, a fine-grained conception of event individuation accepts 3 and denies 4, whereas a person with a (more) coarse-grained conception of event individuation sees one action/event described by both 3 and 4 – Socrates's drinking hemlock at dusk. (Note that a contrastive resolution of description- and focus-shifts is convincing only if alternative accounts about shifty causal emphasis prove to be mistaken – e.g. fine-grained accounts of causal relata (Kim 1976, Bennett 1988)).

[73] Another option, briefly mentioned by Schaffer (2012), is to treat 'causes' as ambiguous between causation and the sentential connective specific to causal explanation as suggested by Davidson (1980a: 162). However, 'causes' does not behave like vague or ambiguous notions and does not display the context sensitivity of indexicals, gradable adjectives or other relational terms.

First, 'causes' does not seem to display the lexical ambiguity of terms like 'file' or 'suit.' But note that 'causes' can be ambiguous between 'the cause' and 'a cause,' and some causal claims can be read as causal explanations (e.g. 'Moderate smoking caused A's cancer'). Also note an intuition of scalarity – that something can be more or less of a cause or that that causes take on quantity-like properties that can be added, subtracted or divided – as common sense does associate magnitudes with events: if two fires join and cause harm to a forest, the feeling is that the bigger fire is more of a cause than was the smaller fire even if natural languages fail to host equivalent expressions.

Second, 'causes' does not seem clearly gradable like other context sensitive terms such as 'flat' and 'tall' – for instance, it cannot take degree modifiers ('very', 'quite', 'somewhat', 'slightly', 'thoroughly') or be part of natural comparative constructions ('this is more of a cause than that was' or 'this was a small cause and that was a big cause').

Third, causal statements cannot take the clarification technique usually taken by uncontroversial context-dependent terms such as 'flat' or 'empty' (Hawthorne 2004). If I say that a room is empty and someone points out that it is stilled filled with air molecules, I can further specify my claim as 'I meant empty of objects or persons.' In contrast, if it offends common sense, a statement like 'Caesar's birth caused his death' cannot be further specified but as the equally valid counterfactual 'Had not he been born, Caesar would not have died.'

favoured by Hitchcock (1996), Woodward (2003), Maslen (2004), Menzies (2004, 2007), Schaffer (2005, 2012). Feinberg (1970) hinted at both sources. A brief rejection of (aspects of) the pragmatic view can be found in McGrath (2005) and Menzies (2007). A critical examination of both accounts appears in Schaffer (2012).[74] To illustrate, consider two causal claims:

(1) Caesar's birth caused his death.
(2) Brutus' stabbing caused Caesar's death.

Context sensitivity is understood as a matter of conversational pragmatics if there is a variation in the assertability conditions of causal claims, but not in their truth conditions (which remain invariant). To those who take context sensitivity to be a matter of conversational pragmatics (*the causal invariantists*), Caesar's birth and Brutus' stabbing equally count as causes of Caesar's death, but one of them is selected as 'the cause' on account of our immediate interests and expectations. On this view, 'Caesar's birth caused his death' is a true causal claim, but we do not select it since it is irrelevant or conversationally inappropriate.

Alternatively, context sensitivity is understood as a matter of semantics if there is a variation in the truth conditions of causal claims rather than in their assertability conditions. On this view, 'Caesar's birth caused his death' is false rather than irrelevant. To those who take context sensitivity to be a matter of semantics (*the causal contextualists*), selection is understood as being constitutive to the causal concept itself rather than a subsequent pragmatic consideration. Indeed, context sensitivity seems to be semantic in both a logical sense (the meaning of what counts as a cause is defined through selection principles), and a psychological sense (just as perception operates through selective screening influenced by expectations, attention, or motivation, the meaning of what counts as a cause is intimately connected to our immediate interests, purposes, or background assumptions). Causal contextualists do not claim that

[74] Usually, the discussion of context sensitivity follows the analogy with epistemic contextualism, as the structure of the debates seems similar; nonetheless, there are important differences. Causal contextualism is in a sense more fundamental than epistemic contextualism. Knowledge attributions require a causal explanation of how context features are relevant to these attributions; however, causal claims already seem to demand a contextualist analysis. In addition, there is significantly more data supporting causal contextualism than epistemic contextualism. Finally, their theoretical motivation is different: causal contextualism is not motivated by the need to offer a resolution of sceptical puzzles (DeRose 1995), an explanation of concessive attributions (Rysiew 2001), or attractive solutions to the lottery problem (Cohen 1998) or the Gettier problem (Lewis 1996).

pragmatic considerations are not present in the causal discourse, but deny their potential to offer a complete explanation of (all types of) context sensitivity.

2.3 A Mixed View?

None of these options is problem free. Pragmatics cannot fully handle all cases of context dependence, as it cannot explain our willingness to assert negations,[75] or why causal claims do not comfortably pass the cancellation test usually passed by conversational implicatures (a technical term coined by Grice to refer to what is suggested by an utterance).[76] These points may tip the balance towards semantics, but context sensitivity appears to be difficult to locate in the semantic machinery (Schaffer 2012) (cf. Chapter 1, Section 2.2). The remaining options are to leave the matter undecided or argue for a mixed view. There are at least three interpretations of a mixed view:

(1) Context dependence is sometimes semantic, sometimes pragmatic, depending on what instance of context dependence one is focused upon. For instance, it might be that some data from causal selection, causal inquiry and multiple alternatives is explained pragmatically, whereas the nearby sentential sensitivity ('rather than' clauses, event specifications, focus shifts) is semantic. As the data is complex, the distinction is not meant to be clear-cut and no obvious criterion for a differential treatment is evident: focus shifts or 'rather than' clauses can also be pragmatically explained, while causal selection can also be a matter of semantics.

(2) Whether context dependence is a matter of semantics or pragmatics will eventually turn on our intuitions. 'Caesar's birth caused his death' may be considered by many

[75] Pragmatics cannot explain our willingness to assert negations, e.g. 'The fact that the table has been laid down for dinner with knives on it does *not* cause the death of that person – what caused it was the stabbing'. In this case, Gricean maxims will explain only our rejection of this fact as conversationally unacceptable, but not our tendency to assert negations (McGrath 2005, Schaffer 2012). Asserting negations seems rather linked to an intuition of falsehood.

[76] Causal claims do not pass the cancellation test (usually passed by conversational implicatures). For instance, if I say that coffee keeps me awake I can cancel the implicature that I do not want some by saying 'although I don't want to say I wouldn't have a cup'; similarly, I can cancel the thought that I want a cup of coffee by saying 'although I don't want to say I would have a cup'. Not so for the causal claims, as one cannot say, e.g., 'The way the table was arranged for dinner with knives on it caused the death of that person, although I do not want to deny that the stabbing might have had a role as well', or 'Socrates's drinking hemlock AT DUSK caused his death, although I do not want to say that the time when Socrates drank hemlock made any difference to his death'. Pragmatic attempts to cancel the initial meaning cannot save the acceptability of causal claims (Schaffer 2012).

true in a sense but irrelevant to the conversational point. To others, it is undeniably false. The force of a particular intuition can be fostered by different framings: if a house catches fire, no scientist will take the presence of oxygen to introduce a completely false causal judgment. Similarly, morally loaded cases have a substantial impact on what causal statements we take to be true or irrelevant.

(3) Whether context dependence is a matter of semantics or pragmatics will eventually turn on how we interpret the concept of *causal field*. This notion was introduced by Anderson (1938) as a restriction on the Millian network model so that we could make sense of singular causal claims[77] (cf. Chapter I), and was further developed by Mackie (1965, 1980):

> The cause is required to differentiate, within a wider region in which the effect sometimes occurs and sometimes does not, the sub-region in which it occurs: this wider region is the causal field. (…) This modification enables us to deal with the well-known difficulty that it is impossible, without including in the cause the whole environment, the whole prior state of the universe (and so excluding any likelihood of repetition), to find a genuinely sufficient condition, one which is "by itself, adequate to secure the effect." (Mackie 1965: 249-250).

Singular causal statements of the form 'C causes E' are elliptical and need to be expanded into 'C causes E relative to a causal field F', where F refers to an assumption or a set of background assumptions made about singular causal judgments. If, for example, the causal field proves to be a notion assumed for conversational purposes, context sensitivity will be a matter of pragmatics; alternatively, if causal fields are supposed to fix the truth of causal claims, then context sensitivity will be a matter of semantics.

Taking a step back, the goal of Section 2 was to illustrate the context sensitivity of causal claims and discuss its sources. On the one hand, it emerges that the context sensitivity of causal discourse is widespread, uncontroversial, and supported by significant data, far greater than the data supporting the context sensitivity of knowledge ascriptions. On the other hand, although causal language shows indisputable

[77] We have a Millian image of events as caused by a multiplicity of preceding and simultaneous conditions and represent the entire causal history of the universe as vast and complex network. According to the network model, every event is necessitated by the preceding causal structure and the universe keeps flowing incessantly, with no sudden breaks or fresh starters. What might have seemed 'exciting, chancy, creative, developing, or unpredictable' is but a mere succession of sufficient causes following uniform principles (Steward 1997, 2008). However, we take the world to contain specific causal sequences (to be marked by unexpected breaks and new starters). To remove the difficulties in Mill's theory and make sense of particular causal sequences, Anderson (1938) introduced the idea that singular causal claims are made in a determined causal field, a notion meant to capture the characteristic circumstances in which a causal sequence occurs. For instance, the question 'What causes influenza?' is indeterminate and needs to be further specified as 'What causes influenza in (some) human beings?'.

pragmatic sensitivities, not all context sensitivity is pragmatic and it is not yet clear how to implement semantic sensitivity in a plausible semantic framework (cf. Chapter I, Section 2.2). So far, Sections 1 and 2 substantiate the basic premises of the initial argument:

P1) Moral responsibility claims rest on causal claims: if an agent A is morally responsible for an outcome O, A must have performed an act (action and omission) that caused O.
P2) Causal claims of the form 'C caused E' or involving similar causal verbs are context sensitive, that is, their truth or acceptability varies across contexts.

The obvious question is then whether moral evaluation inherits the context sensitivity of causal claims.

3. Moral Lessons

3.1 Examples

To give more substance to the idea of context sensitivity, consider the following examples:

(E1) In Sydney some time ago a motor cyclist was exceeding the speed limit; a traffic policeman, also on a motor cycle, chased him, and soon they were both traveling, according to the reports, at 70 m.p.h. Then an unobservant citizen stepped off a bus into the policeman's path; in the crash that resulted the other man was killed at once; the policeman died next day. There was some disagreement as to who was responsible for this accident. The police announced that when they caught the original speedster they would charge him with causing the two deaths. The general public was inclined at first to hold the policeman responsible for the other man's death, but tended to change its mind a little when he died himself. So far as I know, no one said that the man who stepped off the bus was to blame for his own death and the policeman's, but this is a view that could conceivably be held. In addition to these three simple answers to the question "Who was responsible?" there are several less obvious or more complex ones--for example, that no one was responsible, that some unmentioned person or persons were responsible, that the responsibility was shared, or that perhaps even apart from such sharing someone was not fully but only partly responsible, and so on. (Mackie 1955: 143).

(E2) [S]uppose that a teen-age boy, after enduring years of various hardships and horrors, kills one of his two abusive parents. Let's say it is the father. Is the boy morally responsible for the father's death? Typically, the events leading up to this sort of thing will be varied and complex. They will include, among many other things, abuse by the parents, wilful ignorance by family and friends, failures by the school and social service systems, some more immediate precipitating event, and finally decisions and actions by the boy within this broader context. Now consider: according to a plausible account of moral responsibility, the boy is morally responsible for his father's death only if he is causally responsible for his father's death. (…) [T]he explanatory salience that a causal contributor has depends (in part) on the interests and purposes operative in the context of explanation. Finally, consider that the interests and purposes operative in relevant contexts will be varied. Citing only a few (…): those operative for the police at the scene of the crime; those operative for the judge during the trial; those operative for the judge during sentencing; those operative in the evaluation of the social service agencies involved; those operative for the social workers and psychologists charged with devising a treatment plan for the boy; those operative for the boy's priest during the boy's confession; etc. (Greco 2008: 435-36).

It is important to note that attention to the context sensitivity of causal claims brings an awareness of the complexity of causal determinants anteceding an outcome of interest. To get clear on the causal question 'What caused this outcome and why?,' the usual strategy is to divide the causal map of an occurrence into different causal contexts, each characterized by an expected course of events. Answering the causal question depends on the focus of inquiry and on what we take to be the reasonable or expected course of events. The answer to the question 'What caused this outcome and why?' will *vary* with the expectations implicitly assumed when formulating the causal question. In (E1) for instance, taking a legalistic perspective and assuming that it is expected for the police to pursue speedsters or for people to get off buses without fearing speeding motorcycles, but that it is not legal for cyclists to speed carelessly, it is the speedster who ultimately caused the death of the unobservant citizen. Alternatively, taking a less legalistic viewpoint, one might assume that policemen should not pursue speedsters regardless of other citizens' safety and conclude that the policeman caused the death of the unobservant citizen. Even the unobservant citizen may be held responsible in a world with slightly different speeding regulations and street behaviour rules. So assuming different causal contexts leads to offering different answers to the main causal question.

Section 1 defended the view that explaining responsibility attributions typically includes the specification of the relevant causal connection between agents and outcomes, which takes the form of a causal claim. However, we have seen that what counts as the cause of a certain event (and what in turn justifies to some degree responsibility attributions) is context sensitive (Section 2). To the extent to which responsibility ascriptions rest on causal claims, there is a legitimate question whether they inherit or not the context sensitivity of causal discourse. Here one could take one of the following options:

(1) Treat responsibility as not inheriting context sensitivity from causation because claims of responsibility are evaluated by quantifying over contexts.
(2) Treat responsibility as not inheriting context sensitivity from causation by treating attributions of responsibility as fixing a certain type of context.
(3) Treat responsibility as inheriting context sensitivity from causation.

These views might work as follows:

(i) "Agent A is responsible for outcome O" iff "An action of A's caused O" is true in some/a few/most/all contexts.

(ii) "Agent A is responsible for outcome O" iff "An action of A's caused O" is true in the specified context C*.

(iii) "Agent A is responsible for outcome O" is true in context C iff "An action of A's caused O" is true in context C.

In different terms, the options are a) to assess moral responsibility by considering all relevant causal contexts (a bird's eye perspective over the causal complexity preceding an outcome of interest); b) to argue that we are able to focus attention on a particular context as *the* relevant context; c) to admit that context fixing is not possible and that moral responsibility claims follow the contextual variation of causal claims.

3.2 Evaluating the Options

(1) The option of evaluating claims of responsibility by quantifying over contexts suggests either that a certain causal claim will continue to be true in some/all contexts and will end up underpinning the final moral evaluation, or that moral assessment can somehow be derived from the equal and detailed considerations of some/all causal contexts, despite their apparent incompatibility. Given the causal complexity preceding an occurrence of interest, it is unlikely that a single causal claim will continue to be true in more contexts – unless that causal claim is the more general 'All causal antecedents caused the occurrence of interest', which is true but uninformative. On the other hand, if moral assessment is based on a careful consideration of most or all relevant contexts, arriving at a clear responsibility evaluation would require measuring the causal involvement of each contributing factor (e.g., in E1 one would need to assess the causal weight of the speedster, the policeman, the unobservant citizen, and perhaps many other relevant factors). However, it is likely that an assessment of causal magnitude would be inaccurate and subjective (cf. Chapter III, Section 3.2). Lastly, if the moral evaluation considering most of all relevant context appeals to the purpose of inquiry and the criteria governing it, the solution collapses in the second option, *viz*. offering reasons of a pragmatic or semantic nature.

(2) According to the second option, a certain type of causal context can be fixed such that no variation in responsibility ascriptions occurs. Those who take the context sensitivity of causal claims to be pragmatic appeal to relevance rules to fix the adequate context: it is Brutus's stabbing that causes Caesar's death, as claims such as 'Caesar's birth caused his death' are simply irrelevant to the conversational point. However, beside the fact that not all context sensitivity of causal claims is pragmatic, in more complex cases there is no clear notion of causal relevance: although one might blame the original speedster in (E1) above, one cannot hold that the policeman or the unobservant citizen have no causal relevance whatsoever. Along the same lines, perhaps Caesar's birth is not *completely* irrelevant to causing his death.

In more complex cases, one may fix more contexts, i.e., moral responsibility could be ascribed by considering *causal paths* leading to outcomes, where a causal path is marked by 'good representatives', i.e. landmark events relevant to a certain outcome (Swanson 2010). According to this view, the unobservant citizen's birth is not a good representative on the path leading to his death, whereas, say, the motorcyclist's exceeding the speed limit is such a representative. Responsibility is evaluated by taking into account a series of relevant events on the causal path towards an outcome – that is, we apply a selective pragmatics by quantifying over the relevant contexts describing the causal path to the outcome. This view seems better because it takes causal complexity into account. However, there are no distinctly expressed criteria related to the number of good representatives or to what makes a good representative really good or really representative. The criterion cannot be that, holding other central aspects of the causal path fixed, a good representative would have made a significant difference to the outcome, as most events on the causal path are of this kind. Second, path-pragmatics explains the exclusion of irrelevant causes (why only some causes are counted as representatives on the causal path), but it does not point out a way out of disagreement in complex cases like (E1) or (E2). After all, 'come up with good representatives' is not a rule that guarantees impeccable causal paths that everybody agrees with. If what is causally relevant is going to be a contentious matter, there is no clear way of specifying relevant contexts or events on the causal path.

Nonetheless, there is another way to fix a context. This option takes the context sensitivity of causal claims to be a matter of semantics, a feature of our causal concept, rather than a subsequent pragmatic consideration. Consider how legal methodology and practice determines the relevant causal context for the assessment of responsibility.

Given that what matters to the legal system is to secure compliance to certain standards of conduct, causal judgments in the law compare the actual (and wrongful) course of events to an alternative situation in which the defendant acts as the law requires (Schaffer 2010). Simply put, the law is not interested in all causation, but in a specific context, *viz*. the context in which the defendant acts wrongfully. Legal ascriptions of causation fix a certain type of contrast for the cause (namely, lawful conduct on the part of the defendant), and a certain type of contrast for the effect (namely, an outcome in which the plaintiff meets a better fate). In (E1), the law specifies the relevant context through this contrast schema – e.g., that the speedster's acting as the complaint specifies *rather than* acting lawfully caused the unobservant citizen to die *rather than* survive. In different terms, it is the speedster who is responsible for the death of the unobservant citizen because his conduct was not lawful. Causal judgments in the law could thus be defined as contrastive judgments focused on a particularly determined scheme.

(3) The third option is to admit that context fixing is unlikely and that moral responsibility claims follow the contextual variation of causal claims. If the target is *moral* rather than *legal* responsibility, one must observe that moral theory lacks the resources available to legal practice, i.e. the means to facilitate the regimentation of complex situations. Moral theories operate with a more nuanced or sensitive set of ideas about responsibility and in this sense they are more open or attentive to the details of a situation or to the various ways an agent can be morally responsible. This does not mean that legal criteria of responsibility are wrong, but that they are less nuanced, mostly because the law needs to arrive at clear verdicts instead of leaving matters open for further, perhaps endless debates. For these reasons, moral responsibility claims are more exposed to relativistic worries. Discussing how practical interests guide causal emphasis in the law and distinguishing them from ordinary moral responsibility claims, Wright remarks:

> (…) [T]he determinations involved in establishing that a certain factor not only was a cause but also should be held responsible as "the" cause often are inseparable in the conversations and minds of ordinary people. The distinction may not be explicit in conversations, but it is certainly implicit. It is essential to distinguish each step in the adjudication of legal disputes to ensure that the causal inquiry will focus *only on the legally relevant potential causes* (…) (Wright 1985: 1744, my emphasis).

The law keeps the causal inquiry focused by using 'contextualising and doctrinal filtering devices'[78] to achieve a level of specificity that serves its purposes, i.e. to fix what is legally relevant. Because legal methodology can individuate the elements involved in the causation of an outcome, identifying and ascribing legal responsibility appears to be in principle less demanding. Of course, it does not mean that legal theory has fully objective and infallible procedures for how contexts determine relevant contrasts *in every case*, but it allows us to prove certain causal facts and get a handle on certain matters. In contrast, moral theories lack a developed inventory of context-fixing resources so as to manage causal complexity and settle on particular morally relevant contexts. For that reason, moral responsibility claims seem more exposed to inheriting the context sensitivity of causal claims, a predicament which motivates a version of moral contextualism. This version of moral contextualism (a causation-based contextualism for moral responsibility claims) holds that a variation in the assessment of causal sequences may entail a variation in the assessment of moral responsibility. As such, it needs to be distinguished from moral contextualism understood as a view about the justification of moral beliefs (parallel to epistemic contextualism).[79] I will next make a few suggestions about how such an idea might be motivated and defended.

3.3 Moral Contextualism

Consider the following thesis:

[DILEMMA]: In complex situations such as those exemplified by (E1) and (E2), there is either no relevant causal context for determining moral responsibility, or there is always a (really) relevant causal context for determining who is genuinely morally responsible.

---

[78] For instance, in the tort of negligence, these devices specify the actionable damage, the proximate cause, the duty of care, the breach of duty, etc. (Stapleton 2009). In her 2008, Stapleton expresses this idea more clearly: '[I]n contrast to metaphysics, the conceptual framework and methodology of the Law provide filtering devices: that specify a small finite number of factors whose possible involvement in the existence of a particular phenomenon is subject to investigation; that pinpoint which particular phenomenon in the actual world is being examined; that specify relevant hypothetical comparator worlds; and that, within both the actual world and relevant hypothetical worlds, individuate the factor and the phenomenon of interest given the purposes of the legal inquiry'. (Stapleton 2008: 448).

[79] Baumann offers a quick chart of the different forms of moral contextualism: 'Dreier (1990), Norcross (2005a), Brogaard (2003, 2008), Greco (2008), Wedgwood (2006b), Montminy (2007) and Jenkins and Nolan (2010) propose contextualism for core moral terms (see also Unger 1995 and critically: Weatherson 2008, Sect. 4 and Schroeder 2009, 284-287). There is also growing debate on the context sensitivity of normative ought-claims: (see e.g., Bjoernsson and Finlay 2010, esp. the first two sections and Wedgwood 2006a, 151-152). Hawthorne (2001) and Rieber (2006) have tried to apply attributor contextualism to the concept of a free action while Norcross (2005b) defends a contextualist analysis of ''harm''. Sinnott-Armstrong (2006a) defends contrastivism with respect to justified moral belief but does not hold a contextualist version of contrastivism'. (Baumann 2011: 219, fn. 21).

The first horn indicates that we have no reason to deem one context as more relevant than another. All are relevant, and as a result, responsibility is shared amongst *n* factors. The second horn indicates that complex situations such as (E1) always have a relevant cause and a genuine answer to responsibility questions (an authentic explanation) – e.g., the speedster is actually at fault and genuinely responsible for everything. Admitting the first horn is difficult to square with our moral responsibility practices – we do not stop at blaming the whole of antecedent conditions leading to an event. Admitting the second horn is difficult to square with the reality of disagreement. Alternatively, one may accept the second horn but deny we could ever know what the truly relevant context is – but this view also goes against moral practices, as we often take responsibility ascriptions to be justified. In contrast, moral contextualism takes a middle way. There are reasons to choose one context rather than another (against the first horn), but they are connected to interests, purposes, or background assumptions (against the second horn). In different terms, there is always a certain relevant context, but never a really privileged one, objective, or independent from our aims and expectations. According to this view, the truth or acceptability of moral responsibility claims varies with the perspective from which causal judgments are made. Responsibility ascriptions seem thus to conform to our inclinations (which seems to explain the different views and the subsequent shift in attitudes of those assessing E1).

Moral contextualism invites optimistic or pessimistic readings. Optimistically, one could hold that selection is guided by a type of benign interest-dependence which does not affect negatively moral claims (Greco 2008). It is plausible to think that practical purposes and interests of individuals overlap to a large extent, which creates stability across contexts and helps the selection of the relevant ones. Indeed, it would be alarming to think that the acceptability of moral responsibility ascriptions is randomly dependent on our caprices. For that reason, convincing moral theories limit the ways our subjective interests and purposes impact the acceptability of moral judgments, and in this sense, moral theory could opt for a version of contextualism without fearing relativism or moral nihilism. Pessimistically, this recommendation must be taken *cum grano salis*. After all, our interests are more variable than has been recognized to date. The intuition of an apparently stable and benign interest-dependence is haunted by the presence of deep theoretical and practical disagreements. Moreover, we still do not have an account of what is actually a normal, fully justified, or reasonable

interest, aim, or expectance. Finally, there is a disheartening image of responsibility judgments coming from moral psychology – they seem to vary in different cultures, to be influenced by our understanding of norms (Hitchcock and Knobe 2009), affected by framing effects (Sinnott-Armstrong 2006b), or to be rather variable emotional responses rationalized post-hoc (Haidt and Bjorklund 2006).

Two objections need noting:

(1) An often voiced worry is that such a position is too relativist-friendly and cannot avoid collapsing into a form of radical scepticism. This objection is undoubtedly fair, but a possible answer would be that the task of moral contextualism is not to specify relevant contexts, but to show how to interpret causal and responsibility claims, how to delineate the causal map of a complex case, and what the conditions required for context-fixing are. In different terms, the context sensitivity of causal claims demands from us to be explicit about focus and the way focus can change. Insisting on a *methodological* point, moral contextualism holds that causal and responsibility claims do not make sense alone but relative to features encoded in the context. Defenders of this view will only need to show that such contexts are assumed by speakers in virtue of their interests, aims, or background assumptions. Accusing moral contextualists of failing to specify 'which contrasts, whose context, and why' (Sinnott-Armstrong 2008: 453) erroneously assumes that they must favour particular contexts over others, but there is nothing preferential in their doctrine.

(2) Another worry replicates an objection voiced by the critics of epistemic contextualism: ordinary agents seem to have a non-contextualist view of what they are doing when they make knowledge attributions. But taking contextualism seriously would entail that agents are not aware how their interests, expectations, or other psychological factors influence their knowledge ascriptions, i.e., it will lead to the implausible idea that agents are blind to the semantic workings of the concept of knowledge. Along the same lines, a contextualist view about causation will attribute semantic blindness[80] to agents when they make causal judgments and hold that they err in the application of the term 'cause' (which seems implausible).

Even if it appears as an objectionable feature of causal contextualism, we do seem to attribute *some* (semantic) blindness. For instance, if X mentions the driver's cornering too fast as the cause of the accident, Y the defective road surface, and Z the

[80] The term was coined by Hawthorne (2004: 107).

defective design of the car,[81] and if it eventually turns out that X was driving negligently, then Y's and Z's judgments were wrong even if they genuinely thought they were right. After all, we are often mistaken about the causes of things. In addition, if particular interests are made clear (suppose that Y is a county surveyor and Z a motor manufacturer), those disagreeing may realise that their surface incompatibility is in a sense justified. (The same thing happens if the case gets tried in a court of law: the defence will stress the defective road surface and design of the car, whereas the prosecution will emphasize the negligence of the driver). But is this enough to show that a causal contextualist is committed to attributing semantic blindness to the speakers? Perhaps not, as it is not clear that we would say that X, Y and Z are necessarily disagreeing rather than that their claims are in a sense compatible. But compatible in which sense? Three remarks are in order:

a) First, certain care is needed when defining compatibility, as it might be held that causal claims such as 'The birth of the driver caused the accident' are on a par and compatible with the claims of X, Y, or Z above: if all such factors contribute to the outcome and are compatible in this sense, then why not include all factors in the past light cone of the event we are interested in? (e.g., the moments immediately following the Big Bang). In different terms, compatibility must be defined so as to exclude Mill-type uninformative claims – 'the world is complex and everything is causally relevant to an outcome O' – or other irrelevant claims.

b) Second, it is fair to say that an optimistic view emphasizes agreement while a pessimistic view stresses disagreement. The causal contextualist might be happy to admit that some semantic blindness is attributed to agents only when they genuinely believe their causal judgments to be incompatible, but not otherwise.

c) Finally, while the speakers might be aware of certain semantic factors related to their use of the term 'causes' (that is, they do not err in the application of the term) – they may still be unaware that the truth value of their causal claims varies. In this sense, the speakers may be blind to a variation in the truth of causal claims, but they are not blind to the meaning of 'causes.'

To summarise the discussion so far, the goal of Section 3 was to evaluate the consequences deriving from the initial premises – briefly, that moral responsibility claims rest on causal claims (P1) and that causal claims are context sensitive (P2). I

[81] Example adapted from Collingwood (1937).

have considered three options – two in which responsibility claims do not inherit the context sensitivity of causal claims because claims of responsibility are evaluated by quantifying over contexts or by fixing a certain type of context, and one in which responsibility claims inherit the context sensitivity of causal claims. I have argued that it is plausible to admit that moral responsibility claims follow the contextual variation of causal claims because moral theories lack contextualising and doctrinal filtering devices as means for the regimentation of complex situations – which leads to a form of moral contextualism. Assuming that a version of moral contextualism about responsibility ascriptions might follow from accepting both **SV** and the context sensitivity of causal claims, what can be done to stabilize causal thinking in moral contexts?

4. Stabilizing Causal Thinking

4.1 Reverse-Engineering Legal Devices

Faced with a sceptical answer to the problem of context-fixing, any positive, constructive answer to scepticism requires an argument that points out criteria of context selection (even if it is difficult to expect that a principle or set of standards will handle all context sensitivity, one could at least hope that such devices will help reduce indeterminacy). At this point, a suggestion would be to learn from legal practice: legal principles defining causation in tort or criminal law could be used to stabilize causal thinking in moral contexts and help moral theory elaborate a robust analysis of responsibility. In principle, this move is conceivable if we agree that law is 'just as likely to influence causal thinking in the moral domain as to be influenced by it' (Cane 2002: 141), i.e. if law and morality are *symbiotic* normative systems (Raz 1982, Hart 1961, Robinson and Darley 1997, Cane 2002). If law and morality are indeed similar normative systems and moral theory implicitly draws on factors that are explicitly codified in the law, then we could treat the moral domain in the same way we treat the legal domain. For instance, we could adopt a contrastivist view on causation in the moral domain, similar to legal contrastivism. As mentioned above (Section 3.2), a contrastivist view of causation in the law holds that legal reasoning specifies a relevant context by fixing a certain type of contrast for the cause (namely, lawful conduct on the part of the defendant), and a certain type of contrast for the effect (namely, an outcome

in which the plaintiff meets a better fate) (Schaffer 2010).[82] More precisely, the law contrasts the actual breach of a duty (which causes harmful effects) with a situation in which the defendant acts lawfully with respect to that (legally defined) duty. Along the same lines, morality could achieve a corresponding degree of determination by contrasting the actual conduct of a person with what is deemed as *permissible* conduct and by identifying the right or better outcomes from a *moral* point of view as the effectual contrast. After all, legal responsibility often tracks moral responsibility. Whilst this is an excellent solution, it is not without limitations. I will briefly mention two.

4.2 Limitations

(i) It is often the case that moral questions do not have a clear answer (an indeterminacy which indirectly offers support to moral contextualism). For that reason, what is permissible may not reflect what is moral and may even go against it. For instance, if a commercial plane is hijacked and it is assumed that hijackers will use it in a terrorist attack, the right moral action could be either to shoot down the aircraft (on a consequentialist approach) or to avoid the killing of innocent passengers on board (on a deontological approach). In this situation, the law could define what counts as the right legal action, but morally speaking any definition would be controversial. In January 2005, a federal provision was passed in Germany (the *Luftsicherheitgesetz*) allowing armed intervention on hijacked planes in order to avoid potential terrorist attacks similar to 9/11. A year later, in February 2006, the same Federal Constitutional Court declared such actions unconstitutional[83] on clear deontological grounds (mentioning human dignity). If these situations have no straightforward moral answer, legal responsibility for shooting down a hijacked aircraft will be morally controversial because it will be attributed whimsically according to whatever criteria happen to define lawful conduct (as the causal contrast) and the better outcome (as the effectual contrast). At this point, one could embrace the idea that morality is exhausted by what we define

[82] Causal reasoning in the law takes the following form: C [breach of duty] rather than C* [lawful conduct] causes E [damage] rather than E* [better outcome].

[83] 'The armed forces' authorisation pursuant to §14.3 of the Aviation Security Act (Luftsicherheitsgesetz – LuftSiG) to shoot down by the direct use of armed force an aircraft that is intended to be used against human lives is incompatible with the right to life under Article 2.2 sentence 1 of the Basic Law in conjunction with the guarantee of human dignity under Article 1.1 of the Basic Law to the extent that it affects persons on board the aircraft who are not participants in the crime.' (Judgment of the First Senate of 15 February 2006-1 BvR 357/05), available at http://www.bundesverfassungsgericht.de/entscheidungen/rs20060215_1bvr035705en.html

as permissible behaviour in various cultures or circumstances; however, many will disagree with such a relativist conclusion.

In addition, attitudes towards the permissibility of certain actions are variable. For instance, considering the permissibility of causing animal suffering, one may observe the differences between vegetarians, vegans, pescatarians, pollo-pescetarians, flexitarians, fruitarians, lactarians, eggtarians, vs. the basic carnivores. In a different example, Nisbett and Cohen (1996) argued that the culture of honour persisting in the American South and affecting people's attitudes and emotions generates an important disagreement between northerners and southerners regarding the permissibility of interpersonal violence.

(ii) It is often the case that causal/moral judgments in the law differ from common causal/moral judgments, and it is not clear that the latter should give way (and even if they give way sometimes, it is not clear that they should do so as a rule). While it helps us to understand moral responsibility ascriptions, the law cannot be invariably trusted to settle standards of moral judgment. Consider the following points:

Reasoning. There is a significant analogy between normative reasoning in law and morality, but closely following the criteria of reasonable reflection (pertinent information, psychological capability, impartiality, or sincerity) does not guarantee corresponding causal judgments. It is true that sometimes judicial reasoning *just is* moral reasoning, but the question arises *when* this convergence is achieved. When deciding particular issues, legal practice is consistent with its previous decisions in similar cases. However, sometimes it departs from them and creates precedents. Given law's authoritative standards, it is often difficult to separate the formal, authority- and consistency-based reasons from the moral, substantive reasons in interpreting causation or making a certain responsibility attribution.[84]

Disagreement. The argument that the law could help us solve moral disagreements and answer decisively controversial moral questions seems to get things backwards. It is precisely because radical moral disagreements persist that an objective and definite

[84] Additionally, even if both forms of reasoning have the same target and follow similar criteria, it does not mean they proceed along the same lines. It is not obvious whether moral judgment is a question of consistently applying rules to particular cases. It is not transparent when morality needs to create precedents and change its evaluative and decisional direction. It is not even clear *how* it makes progress, given that moral practices are deep-rooted and take considerable efforts and significant periods of time to change.

answer to moral questions cannot be clearly pointed out (or cannot be claimed to exist). In addition, if legal principles offer practical solutions to decision problems, the practical point and the principled moral solution may come apart.[85]

Guidance. The law could provide a 'moral exemplar' to folk morality. However, the law is only rarely and peripherally such a moral exemplar. For instance, capital punishment is not rejected everywhere in the US, laws of anti-discrimination often do not affect people's deep evaluative stance or attitudes (e.g., racial, territorial, historical),[86] and environmental law offers us only remote and general reasons to behave morally (we may all agree that pollution is bad, but it may be that no individual has a rational interest in paying additionally for better pollution control equipment). For that reason, the additional thought that law's institutional resources may serve to entrench certain moral practices is even more sobering.

In short, the similarities between law and morality support the idea that law can help stabilize causal thinking in the moral domain, while their differences tip the balance in favour of moral contextualism.[87] It follows that moral contrastivism as a result of reverse-engineering legal devices represents only one of the theoretical alternatives available. The other option is to embrace the mild scepticism entailed by moral contextualism.

5. Sceptical Solutions

Moral contextualism is a sceptical view and taking context sensitivity seriously motivates a version of scepticism about moral responsibility. But why? After all, the fact that we can switch between contexts does not itself entail any form of scepticism – it is not our flexibility or our ability to see things from different perspectives that entails scepticism.[88] Moreover, speakers are sensitive to the features of conversation fixing the context and so are in the position to form reliable judgments about the truth value of causal claims. Naturally, the truth of these judgments is relative to context, so the view

[85] In reply, one could claim that practical answers are to be preferred to solutions of principle, or that lacking a practical point may often amount to 'missing the point' entirely (Cane [2002], for instance, takes this line of argument). This could prove to be true; however, it would define the moral project in terms that many would not be ready to agree with.

[86] Cane (2002: 14, fn. 50) mentions the abortion law in the US, with such polarised views that legal reconciliation seems highly unlikely.

[87] One might have to think further about whether we could get *inconsistent* ascriptions of moral responsibility (i.e., P is responsible for X and P is not responsible for X) or simply *multiple but consistent* answers (i.e., P is responsible for X, and so is Q, and so is R, etc.).

[88] Although it is also possible to argue that perspectivism entails scepticism.

is in this sense *relativist*, but this does not entail that all causal ascriptions are false or that responsibility is impossible (full-blown scepticism).

Indeed, moral contextualism entails a rather mild form of scepticism: it does not hold that moral responsibility is impossible, but that ultimately it is very difficult to identify with confidence. Similar to other sceptical views (Strawson 1994, Rosen 2004, Sinnott-Armstrong 2006a), it is independent of whether determinism is true or not. Unlike these sceptical views, it does not claim that we cannot be morally responsible, in a fundamental sense, because of the way we are (our cognitive and social make-up). It claims that we cannot be ultimately morally responsible because of the complexity of the world and because of the way our causal knowledge is structured. I mentioned that attention to the context sensitivity of causal claims brings an awareness of the complexity of causal determinants preceding an outcome of interest. But what emerges is that the more we know about the causal complexity preceding a particular event, the more uncertain we become about what actually causes it. The more we become aware of the minute specificity of the causal world, the more we begin to realise how difficult it is to explain things fully. As a result, there seems to be a direct relationship between complexity and uncertainty, with the latter beginning to undermine the explanatory stability that should back up a correct moral assessment.

And still, one may ask what justifies the further move from contextualism to scepticism. Why not stop at a sort of relativist contextualism which holds that relative to perspective/context A, an agent caused O (and hence could be morally responsible for O), while relative to perspective/context B, the agent did not cause O (and hence could not be morally responsible for O)? Why do we need to infer from this to scepticism about whether the agent really caused O? On the contextualist view, there is nothing more to be said other than that relative to one context she did and relative to another she did not. After all, contextualism should be an answer to scepticism. In reply, one may hold that we start from the assumption that complex situations like those in E1 and E2 have a genuine answer to the moral responsibility question, but unfortunately, that answer cannot be established with confidence. Had we assumed situations like those in E1 and E2 do not have a clear answer, we would have begged the question. Indeed, moral contextualists may be positively satisfied with a relativist resolution, but

as long as they assume an authentic answer to be possible, their position entails the mild form of scepticism described above.[89]

This view has similarities with Baumann's (2011) contextualist answer to the indeterminacy of relevant reference classes applied to actions. Baumann compares the following two cases:

(1) A is carefully driving in his car at 20 mph on deserted Z-street. Suddenly a child appears out of the blue and runs onto the street. A cannot avoid hitting and killing the child.
(2) A is carefully driving in his car at 20 mph in the densely populated X-neighbourhood. Suddenly a child appears out of the blue and runs onto the street. A cannot avoid hitting and killing the child. (Baumann 2011: 207-208).

The tendency is to claim that (1) is a case of bad luck, whereas (2) is a case of morally bad behaviour. But here is an interesting possibility: (1) = (2), the deserted street is actually part of the densely populated neighbourhood. And here is the puzzle: according to which relevant context or reference class shall we judge A's action? Events on the street or events in the neighbourhood? Was it a horrible mistake or sheer bad luck? The answer clearly depends on the relevant context, but there is no way of identifying it and thus no way of knowing whether A acted responsibly or not.[90] Moral contextualism follows. In contrast, the view discussed here is not based on the reference class puzzle, but on the semantic/pragmatic context sensitivity of causal claims in complex situations such as (E1) and (E2).

In addition, the view has affinities with Sinnott-Armstrong's (2006a) Pyrrhonian scepticism about the justification of moral beliefs. Is moral nihilism a relevant alternative to common moral beliefs? To invariantists (to whom a certain context is always relevant), moral nihilism is always relevant. To contextualists, it is only sometimes relevant (as the relevant context varies). Against both, Sinnott-Armstrong's (2006a) sceptical answer is to suspend judgment about which context is relevant. In contrast, the view defended here makes a narrow, causation-based claim

[89] In addition, further motivation may come from the fact that 'who is really responsible for this?' is still a question making sense to most people, even if it is possible to sidestep it in the philosophy class.
[90] Baumann rejects two versions of Reichenbach's (1949) idea that the relevant context is the most proximate one ('the smallest and narrowest reference class') as there is simply no sound reason whatsoever to choose *only one* context or *the smallest/narrowest* context. Two options survive: skepticism about the possibility of identifying the relevant contexts and nihilism, the idea that there is no reason to choose one context rather than others. The first option is difficult to square with the fact that we do ascribe moral responsibility and for that reason is unacceptable; the latter option entails moral contextualism. This seems the only game in town: the truth value of judgments such as 'A is morally responsible for an outcome O' varies with the perspective, practical interests, or psychology of the attributor (the individual who makes the judgment). In this sense, moral contextualism offers a preliminary answer to the puzzle, but not a solution.

about responsibility rather than about the justification of moral beliefs, *viz*. that a version of moral contextualism is entailed by the causal footing of (at least some) moral responsibility statements. However, both views share the mild scepticism mentioned above, *viz*. that responsibility is not impossible, but difficult to identify with confidence. In Sinnott-Armstrong's words,

> [This scepticism is not] the dogmatic claim that nobody is justified in believing anything (or that nobody knows anything). This is a straw man. No real sceptic has held such dogmatic scepticism for more than a short while. All real sceptics are Pyrrhonian sceptics like me. (Sinnott-Armstrong 2008: 454).

A final remark: since responsibility may attach to agents, actions, or outcomes, any causation-based and outcome-based account of moral responsibility has to offer some reasons for shifting the emphasis from accounts of responsibility based on *beliefs* rather than choices (e.g. Moya 2006), or on *actions* rather than consequences (Morse 2000, 2004). In need of clarification is also the idea that sceptical intuitions about responsibility may be cognitive illusions because they are insensitive to what the idea of responsibility should actually track (according to recent work in psychology) – namely, the connections between motivation and the objects of responsibility, and the practices of holding people responsible (Björnsson 2010). Discussing these aspects is beyond the scope of the current chapter. However, it is important to mention that the emphasis on outcomes is justified not only by an interest in causation, but also by a practical rationale: outcomes represent the widest class of things associated with responsibility claims. As for the idea that sceptical intuitions are illusory, my intuition is that sceptical insights have their own, perhaps equally important function – but further work needs to be done to understand their relation to responsibility.

Taking a step back, Sections 4 and 5 have shown that the discussion of how context sensitivity of causal claims bears on moral responsibility ascriptions leads to a choice between a) moral contrastivism (Section 4), arrived at by reverse-engineering legal devices, and b) a version of scepticism (Section 5), arrived at because of the difficulties with context-fixing and the idea of permissibility. The latter view seems more plausible given the causal complexity of the world and the differences between law and morality.

6. Conclusions

I have defended the following claims:

1. **SV** If an agent A is morally responsible for an outcome O, A must have performed an act (action, omission) that caused O.

2. It is plausible to claim that the truth or acceptability of moral responsibility claims might inherit the context sensitivity of causal claims and motivate a version of moral contextualism/mild scepticism about ultimate responsibility ascriptions.

3. We might appeal to legal principles defining causation in tort or criminal law to stabilize causal thinking in moral contexts, which motivates a version of moral contrastivism taking *permissible* conduct as the causal contrast and the right or better outcomes from a *moral* point of view as the effectual contrast. This solution, however, has several limitations, which motivate a shift towards the alternative mild scepticism entailed by moral contextualism.

4. An important consequence of taking the context sensitivity of causal claims seriously is that the more we know about the causal complexity preceding a particular outcome (events or states of affairs in the world), the more uncertain we become about what actually causes it (and therefore more hesitant about the right moral assessment).

Context sensitivity is an important feature of causal claims. But connecting it with common sense intuitions about the relation between causation and responsibility has led us to question the identification of ultimate moral responsibility in complex situations. The next chapter takes up a different aspect of causal relations, i.e. cases of causal overdetermination, with a focus on non-reductive accounts of group agency. It explores the concern that group agents would causally overdetermine the effects already caused by their constituent individuals (and thus would morally overdetermine them), but also provides further insight into related issues such as the allocation of responsibility for concerted actions, the weight of contributory causes, and the concept of macro-explanation.

III. Higher-Level Causation

'There is no twisted thought without a twisted molecule.' (Attributed to Ralph Gerard).

Introduction

A number of philosophers have sought to associate agency with groups and describe them as real and robust entities, irreducible to the sets of networked individuals they are constituted of. Recent examples include arguments for group agency realism (List and Pettit 2011), group minds (Pettit 2003), or autonomous collective intentionality (Tollefsen 2002a, 2002b, 2003), but also non-reductive versions of non-summative arguments for 'we-intentions' and joint action (Tuomela 2005, 2007), plural subjects (Gilbert 2009, 2013), or non-distributive collective responsibility (Copp 2006, 2007, List and Pettit 2006, Pettit 2007).[91] Most non-reductive views focus on the criteria required to elevate groups to the status of novel centres of judgment, intention and action, but show comparatively little interest in their actual efficacy as relatively independent entities. In this chapter, I shall argue that non-reductive accounts lack an adequate causal story about how group agents impact the world. While I shall be primarily discussing the problem of causal overdetermination, some light will be shed on related issues such as the idea of macro-explanation, the magnitude of contributory causes, and the allocation of responsibility for concerted actions.

Overview: Section 1 introduces the case for non-reductive views, using List and Pettit's (2011) group agency realism as a working example. Section 2 explores the concern that group agents would causally overdetermine the effects already caused by their constituent individuals. I show that non-reductionists need a coherent causal story independently of whether the overdetermination objection is decisive or not. In particular, I argue that non-reductive manoeuvres generate a degree of commitment to an independent source of causal efficacy and any attempt to reroute it through

[91]Summative views understand groups as aggregates or sums of individuals. Non-summative views are also individualistic but take groups to presuppose further significant interrelations amongst individuals (e.g., shared intentions, collective beliefs, the presence of mutual obligations or expectations). Non-reductive versions of non-summative views or arguments are those inclined to concede groups a certain autonomy. Along the same lines, non-distributive collective responsibility transcends the contributions of individual group members and attaches to groups as self-sufficient moral agents.

individuals in virtue of constitution would entail a breakdown of higher-level performance and effectiveness. Section 3 examines a non-reductionist reply based on the realisation-insensitivity of higher-level causal claims. This line of reasoning is contentious because it is connected to a dependence rather than to a production account of causation (Hall 2004), it tends to conflate explanatory role with real causal efficacy, and it leaves explanation and moral evaluation to bounce fortuitously between levels.[92] Lastly, I discuss the practical value of group agency views – particularly their pivotal role to social explanation, prediction and design – and offer a brief recommendation concerning our stance towards group agency talk (Section 4).

1. Group Agents

1.1 Introducing Group Agents

Common sense frequently associates intentional states and action with certain groups. We say, for instance, that the Roman Senate executed the leaders of the Catilinarian conspiracy, that Britain declared war on Germany in 1939, or that the United Nations promote peace, facilitate social progress and defend human dignity. Group-talk can be interpreted generously, as referring to genuine collective entities, or parsimoniously, as a mere *façon de parler* masking the reference to individual agents. A too generous interpretation may lead to a form of emergentism, with group agency surfacing as a mysterious force, a *Geist* hovering above organisations. A too parsimonious interpretation may lead to eliminativism, which would reduce to individualistic terms any complex organizational properties. Both options seem unsatisfactory, one for its apparent metaphysical extravagance, the other for its inability to fully capture the complexity of the social world. Therefore, a reasonable position taking distance from extremes would admit that reductive individualism is insufficient for an extensive explanation of social practices, norms and institutions, but also concede that 'calling corporations persons creates more confusion and misunderstanding than clarity' (French 1995: 10). Instead, it would be best to claim that certain[93] organized collectives

[92] For instance, think about the responsibility for events such as the UK banking crisis which bounces between regulatory institutions, individuals like Fred Goodwin, and on to others.

[93] The view does not hold for all kinds of groups. The world is populated by a wide variety of collectives, including simple coordinated associations (jazz quartets), integrated complex groups with a formal structure (corporations or governmental agencies), groups with an informal structure (a reading group on Kripke's *Naming and Necessity*), or mere collections lacking any configuration (the group of people at the Old Bookbinders Ale House last week). They all rest on a broad distinction between *aggregates* (less than an organizational minimum) and *conglomerates* (more than an organizational maximum), however, only well-structured collectives are said to have agential properties – e.g., organizations with a formal

are independent 'agents' or 'actors'. As agents, these collectives are a bit less than persons but significantly more than the sum of their individual members, coming to have 'a distinct intentional profile from the profiles of their members'. (Pettit and Schweikard 2006: 33).

A recent example of such an intermediate position is List and Pettit's (2011) group agency realism. They argue that group agents are real and irreducible entities in a straightforward functional[94] rather than ontological sense, but remain faithful to a permissible version of individualism, *viz*. one that allows realistic, non-metaphorical talk about collective entities – a position apparently congruent with the work of French (1984, 1995), Coleman (1974), Rovane (1997), or Braithwaite and Drahos (2000).[95] To defend their view, List and Pettit suggest a range of necessary and sufficient criteria for agency and then argue that groups meet them. To be an agent, an entity must possess:

a) representational states describing how things are in the world,
b) motivational states to stipulate how it wants the world to be, and
c) a capacity to process these states and actively intervene in the world on that basis.

If certain groups fulfil these general conditions of agency, they count as systems acting to realize their motivations according to their representations, i.e. as real agents. Strategy-wise, List and Pettit rely on the difficulty of deriving group-level facts from individual-level facts and on the notion of supervenience[96] as delivering dependence without reduction. While some of these aspects will come up later, a full defence centred on the conditions for agency is complex and cannot be taken up here for reasons of space. Instead, the focus will be the second part of condition c) above – the test of

structure, decision-making procedures, and involving different kinds of roles, hierarchies or task divisions.

[94] That is, agents are not defined in terms of their material properties, but rather in terms of the functions they perform. Note that such an apparently harmless methodological assumption needs justification.

[95] According to List and Pettit (2011) group agency realism occupies a middle ground between the *authorisation theory* and the *animation theory*. Authorisation theory holds that individuals merely authorise group agency talk (Hobbes 1990, Locke 1960, Rousseau 1997). In contrast, animation theory understands group agents as something extra, above and beyond the emergence of coordinated individual dispositions – as argued by Barker (1915) or legal historians such as Otto von Gierke or Frederick Maitland (see Runciman 1997).

[96] The term 'supervenience' could be defined as follows: for two sets of properties A and B, A supervenes on B if there can be no difference in A without a difference in B. A is labelled the *supervenient* set of properties, while B is labelled the *subvenient* set of properties. Reversing supervenience – there can be no difference in B without a difference in A – one arrives to the idea of determination. Supervenience and determination are facets of the same coin.

efficacy – and on the argument that group agents play a causal role that is in a sense irreducible to the sum of individual causal contributions $c_1+c_2+\ldots c_n$ of all members. The test of efficacy is crucial considering the assumption that for an agent to exist is for it to have specific causal powers associated with it. Epiphenomenalism – the idea that agents could exist without being causally efficacious – is not an option contemplated by group agency realism or any other view with non-reductionist leanings, as it would entail a disguised reduction to individualism. If a large part of the literature is dedicated to arguments for extending representational and motivational states to groups (conditions a) and b) above), little systematic attention has been devoted to group agents' actual efficacy as relatively independent entities. I shall argue that non-reductive accounts need a coherent causal story about how group agents impact the world.

While I will use List and Pettit's (2011) group agency realism as a working example, it is useful to briefly taxonomize the extant collectivist views to get a better grasp on what counts as a non-reductive or a non-reductive-oriented account.

1.2 Taxonomy

Current collectivist views are organized along two dimensions, structural and conceptual. On the structural dimension, the starting point is *summative individualism*,[97] which analyses group attitudes ascriptions in terms of the sum of individual attitudes whose content is identical with that ascribed to the group (Ladd 1970, Quinton 1976, Keeley 1981, Dan-Cohen 1986, Miller 2002). According to this view, we cannot fully credit groups with minds of their own, for solely individuals possess them; we cannot assign actions or attitudes to groups, for only persons are entitled to real agency; we cannot hold groups responsible or accountable, because only individuals are traditionally understood as autonomous decision-makers. Ascending structurally, *weak holism* holds that an extensive explanation of social practices, norms and institutions renders summative views insufficient, since coordinated actions seem to require or imply collective attitudes. Weak holists argue that a reductive,

[97] This is the position endorsed, for instance, by Anthony Quinton: 'We do, of course, speak freely of the mental properties and acts of a group in the way we do of individual people. Groups are said to have beliefs, emotions, and attitudes and to take decisions and make promises. But these ways of speaking are plainly metaphorical. To ascribe mental predicates to a group is always an indirect way of ascribing such predicates to its members. With such mental states as beliefs and attitudes, the ascriptions are of what I have called a summative kind. To say that the industrial working class is determined to resist anti-trade union laws is to say that all or most industrial workers are so minded'. (Quinton 1976: 17).

individualist perspective does not adequately describe and explain the complex organizational properties of various groups, and for that reason its popularity decreased with the emergence and evolution of modern organizational theory in the 70s and later. Amongst other things, organizational theory made clear that the relation between group and its members is contingent (members change, organizational identity persists) and the reasons for action specific to an organization are not reducible to the reasons of its individual members. As such, weak holism aims to distance itself from but at the same time safeguard a version of individualism by claiming that group attitudes or actions are individual attitudes or actions with collective content (Bratman 1993, Searle 1995, Tuomela 1995, Kutz 2002). Lastly, *strong holism* attempts to renounce individualist assumptions and holds that groups can be consistently credited with minds, agency, or autonomy in a significant way, different from the way we normally assign these notions to their members (Tollefsen 2002a, 2002b, Pettit 2003). While strong holism can be safely labelled as non-reductive and weak holism as non-reductive-oriented, it is transparent that non-reductive ambitions towards more robust forms of group agency are a matter of degree. On the conceptual dimension, collective notions usually attach to agency, actions, and responsibility, but diversify and particularise in different accounts as discussions about corporate persons (French 1979, 1984), shared agency (Bratman 2009), group agents (List and Pettit 2011), group minds (Pettit 2003), collective intentionality (Tollefsen 2002a, 2002b, 2003), 'we-intentions' and joint action (Tuomela 2005, 2007), plural subjects (Gilbert 2009, 2013), or non-distributive collective responsibility (Copp 2006, 2007, List and Pettit 2006, Pettit 2007, *inter alia*). The emphasis of my discussion will be on strong, non-reductive holism about group agency and responsibility; however, as long as analyses of group minds, shared intentions, plural subjects, or collective responsibility tend towards a sense of autonomy or irreducibility of higher-level notions, they will be included in the extended corpus of views targeted by my argument.

1.3 Higher-Level Causal Efficacy

My contention is that non-reductive accounts of group agency (or accounts with non-reductive aspirations) lack a coherent causal story about how group agents impact the world (that is, how group agents are causally efficacious as relatively independent entities). But why insist that group agents need to play a causal role that is irreducible to the causal efficacy of individual contributions? It might not be entirely obvious that

a defender of holism must hold this. For instance, one might think that a main reason for thinking that group agents exist relates to the worry that their mental states are not simply a sum of what individual members believe or desire, but significantly more, in the sense that collective agents might venture into zones of attitude and action that no member individually follows, understands, or endorses. In reply, we need to keep in mind that a system counts as an agent not only when it displays representational and goal-seeking states and satisfies conditions of rationality (List and Pettit 2006), but also when it acts in the world. In different terms, causal efficacy is a fundamental prerequisite of group agency. Second, at least on a strong holistic view, group agents exist as real and robust entities and for something to exist is for it to have causal powers (paraphrasing Alexander's [1920] dictum[98]). While the intimate connection between existence and causation is controversial, it is safe to claim that no theorist with non-reductive leanings will endorse an epiphenomenalist view or take group agents to be causally inert like abstract or fictional objects (e.g., unicorns, ghosts, sets). Third, as long as such views are non-reductive, their causal efficacy is non-reductive as well, i.e. it is not reducible to the causal efficacy of the networked individuals constituting the groups. In different terms, any non-reductive manoeuvre generates a degree of commitment to an independent source of causal efficacy. Lastly, if outcome responsibility rests on causation (cf. Chapter II), group agents will be morally responsible independently of their members only if they can be causally efficient in an independent sense. It is customary to argue that both groups and their members are legally or morally responsible for different occurrences in the world, but if actual causal efficacy is only associated with individuals, then the notion of group responsibility will need to have a different meaning, perhaps instrumental or metaphorical.

A different, more refined concern is that even in the case of individual agents, it is a very vexed and controversial question whether we can understand agents as acting in addition to the physical material of which they are made up. Therefore, whatever criteria one proposes for group agents, these should not be more demanding than what is accepted as the relevant demand for individual agents. Many people defend the view that one's mental events and states (e.g., decisions) affect the world by being constituted or realised by some physical event which causes physical goings-on. The analogue of this would be the view that the group's decisions affect the world by being constituted

[98] 'To be is to have causal powers' (Alexander 1920: 8).

by something at a lower level, say a set of individual decisions, which realise the group decision. But perhaps one should explicitly talk about and motivate a particular set of claims about individual agency by which to judge group agency. Otherwise, the risk is to bring down the effectiveness of one's arguments by holding group agents to a standard that even individual agents cannot meet. As an answer to this concern – but without getting into details – I am restricting my discussion to the general conditions of agency that any system should meet, i.e. the conditions reflecting a widespread consensus amongst psychologists, economists or philosophers of mind: representational states, motivational states, rationality, and causal efficacy. I am appealing to this widespread consensus to avoid a comprehensive review and defence of every detailed condition of agency, which is obviously beyond the scope of this chapter. Second, I am happy to concede that most of my conclusions about the higher-order causal efficacy of group agents may equally apply to the debates about mental causation (although it is not my intention to discuss the concept of mental causation).[99] Third, it seems somewhat methodologically onerous to start the discussion of group agency or collective action from a detailed motivation and defence of one's commitments regarding mental causation. Considering that, perhaps it makes more sense to organize the discussion around the most general and broadly accepted conditions of group agency.

Finally, higher-level causal efficacy raises important questions about the function of social roles. One natural view is that groups act in virtue of some individual persons with particular social roles acting. For instance, a political party acts by its organising committee acting which in turn happens by its chairman acting. This seems to amount to giving up on group agency. But, an objector might insist, this has not really reduced the social to the non-social, for some particular individual's action only constitutes the group's action if the individual occupies a social role. Some individual person acting does not constitute the group's action unless she is given that social role by the group. Indeed, the distinction between roles and role bearers requires more attention. One of the reasons why we can speak about attitudes and actions at the level of institutions is justified by their being built up from roles, with the role bearers being supposed to act in their institutional or official capacity. So the intentional states and actions of the role

[99] I am saying 'most' of the claims because important differences persist – for instance, the internal constitution of group agents is different from that of an individual.

bearers can be attributed to the roles. Now as far as the general question of causal efficacy is concerned, it seems that only individuals are causally relevant and what changes is the guise under which they act. However, matters are more complicated when it comes to responsibility and obligations, where the distinction between roles and role bearers may be used to defend a certain sense of independent group agency and responsibility.

For instance, if the daughter of a prime minister (PM) is kidnapped by a group who threaten to kill her unless she authorizes the government to release a dangerous prisoner, the PM has a special duty towards her family, but as a government official she has an obligation not to release the prisoner.[100] The PM eventually decides to authorize the release, which seems to generate a case in which the government acts and is responsible, but not the PM (because she is exonerated by the fulfilment of her special obligation towards her daughter). More precisely, the PM is not blameworthy for her action, while her action constitutes[101] the action for which the government is blameworthy (as the government has no excuse). Special duties make a major moral difference, however, the emphasis falls on the nature and the weight of what governments require from their officials. At this point, intuitions struggle. If the PM accepted the office and it was specified that her position *does not require* her to override special duties, then the PM's action of releasing the prisoner does not contradict the government's action – because a government's duties are mirrored in the duties of those occupying the official roles. However, it is hard to believe that the acceptance of the PM office does not imply giving full priority to the interests of the government or its citizens. Alternatively, if the PM accepted the duties and responsibilities associated with the office, and these duties *required* overriding her special duties, then the PM is directly responsible for releasing the prisoner. But again, there must be something morally suspicious about the government which requires its officers to do something that would be wrong for them all things considered – imagine a kamikaze administration whose officers accept with the office the duty to sacrifice themselves and their families (in the event that this is necessary). Finding a balance here is difficult. If the obligations of an institution are reflected in the obligations of those occupying institutional roles, the institution cannot be morally sound if it demands its role fillers

[100] Example adapted from Copp (2006).
[101] I leave aside the technical details regarding the concept of 'constitution', or those related to the notion of (political) 'representation'.

to perform an action they believe to be morally wrong for them. But on the other hand, it could not be a sound institution if its principles were systematically overridden. Without getting into further detail, it is important to note that in both cases the PM's action is bound up with the action of the government and there is no possible description of the case in which the government acts and is morally blameworthy without the PM acting and being morally blameworthy (Ludwig 2007: 421). In general, considering the distinction between roles and role bearers, it is important to note that one becomes willing (and realizes that it is rational) to accept a role only after admitting the theoretical independency of the group. More precisely, one accepts the group-related reasons only after acknowledging the primacy of the group when making decisions, embracing conclusions, and so on. However, accepting its *primacy* does not amount to recognizing its independent *agency*.[102]

Taking a step back, the goal of Section 1 was to introduce the idea of group agency and taxonomize extant collectivist views. I have argued that non-reductive views (and certain non-reductive-oriented accounts) are committed to an independent source of causal efficacy, that is, group agents need to play a causal role that is irreducible to the causal efficacy of individual contributions. In different terms, such views need to test for a mode of action consistent with their non-reductive way of conceiving of agents. In the next sections, I will use a version of strong holism (List and Pettit's [2011] group agency realism) to discuss the issues associated with the idea of higher-order causal efficacy – mostly the problem of causal overdetermination.

2. Overdetermining Causes

2.1 Causal Liberalism and Some Terminology

At least as far as ordinary causal talk goes, groups are causally efficacious. Common sense takes entities from different categories to be causal relata. For instance,

The earthquake caused the tidal waves.

That is was snowing caused us to remain indoors.

Low frequency active sonars caused the whale strandings.

Slope angle and orientation caused the avalanche.

Herostratus caused the destruction of the Temple of Artemis at Ephesus.

[102] For further discussion of the way group agents control for the performance of some individuals (and make their action 'inevitable'), see Section 2.3 in this chapter.

These examples suggest that events, facts, objects, properties, or people can cause or be caused, so the following thesis seems true:

[CAUSAL LIBERALISM]: Causal relata may include entities from different ontological categories.

[CAUSAL LIBERALISM] takes our causal talk at face value and offers a flexible answer to the problem of relata individuation (that is, the question whether relata are events [coarse-grained] or facts, features, tropes, states of affairs, etc. [more fine-grained]).[103] A difficulty entailed by [CAUSAL LIBERALISM] is that in many cases it is natural to speak of an effect as being caused by entities belonging to different ontological levels. For instance, it is possible to claim both that

(1) The Roman Senate executed[104] the leaders of the Catilinarian conspiracy,
(2) The members of the Roman Senate executed the leaders of the Catilinarian conspiracy,
or even that
(3) Cicero executed the leaders of the Catilinarian conspiracy.[105]

This raises the following question: is a particular effect E caused by the group agent, the individuals constituting it, or both? If causal work is ultimately done by individual determiners, the sense in which group agents make a difference in the world requires explanation.

Let me first make some terminological clarifications. Usually associated with aggregates, *joint causation* occurs whenever multiple actual causes c_1, c_2, …, c_n contribute to bring about an effect E and each of the c_1, c_2, …, c_n is necessary for bringing E about. Consider a classic case of accomplice liability: Agent A_1 holds a gun

[103] The idea that causal relata are as diverse as common causal talk suggests (with fact-causation being more fundamental than other types) has been defended by McGrath (2002, unpublished manuscript).
[104] I take 'executed' to be a causative verb. For reasons of space, I leave aside a detailed discussion of the equivalence thesis, *viz*. 'For any causative φ, X φ-ed = X caused a φ-ing to exist' (but see Moore 2009: 5-19). On the equivalence thesis, (1) could be rephrased as 'The Roman Senate **caused** the execution of the leaders of the Catilinarian conspiracy.' The same goes for (2) and (3).
[105] As consul, Cicero delivered four excellent orations against Catilina and convinced the other members of the Senate, including Caesar who proposed less excessive measures, that execution was the right measure.

on the bank clerk, A_2 takes the money from the cash register during a robbery, A_3 drives the escape car. The effect – here a successful robbery – is jointly caused by all acts. *Causal overdetermination* occurs when at least two causes c_1 and c_2 are individually sufficient to bring E about. Associated with non-distributive groups, *constitutive causal overdetermination* occurs when both group agents (macro-causes) and the individual members they are constituted of (micro-causes) are in themselves sufficient to cause an effect E. Two remarks are in order: first, it might be argued that causal overdetermination is merely joint causation underdescribed (Bunzl 1979). However, this is not the case for the instance of constitutive causal overdetermination discussed here: defenders of group agency realism argue for more than the summative view described by joint causation. Second, the constitutive causal overdetermination associated with groups must be differentiated from material constitution (what Mackie [1980] calls quantitative overdetermination) – e.g., when one rock hits a window, the rock's eastern and western hemispheres are overdetermining causes (distinct and individually sufficient) of the window shattering.

It is not clear that, in general, overdetermination should be thought of as detrimental to a theory; however, it is usually taken to be *prima facie* problematic. Assuming that it is *prima facie* problematic, the relevant argument against group agency realism is, in short, that if macro-entities like group agents existed, they would overdetermine any effect they may have on the world, since these effects would also be caused by the individuals groups are formed of. However, since these effects do not seem to be overdetermined, it follows that macro-entities such as group agents do not exist (and group agency realism fails). A similar thesis has been defended by Merricks (2001) with respect to all non-living macroscopic physical objects (mountains, oceans, guitars, computers) and evokes the exclusion problem in the philosophy of mind (Kim 1989, 1993a, 1993b): mental properties have to either be reduced to physical properties or eliminated altogether because they would overdetermine the effects of these physical properties. In contrast, several philosophers have argued that widespread overdetermination – or at least the type of overdetermination brought about by nonreductive metaphysics – is not that problematic (Schaffer 2003, Sider 2003). Admittedly, causal overdetermination may be objectionable or ultimately prove to be not as bad as usually thought; but in both cases there are important lessons to be drawn for group agency realism.

2.2 Overdetermination Is Problematic

Assume first that causal overdetermination is objectionable – say, because it amounts to causal redundancy, or because it appears to be metaphysically confused and epistemically undermotivated (Merricks 2001, Kim 1993a, 1993b). These objections will receive a careful assessment in the next subsection; meanwhile, it is important to note that not all kinds of overdetermination are equally controversial. A distinction is frequently drawn between a) *standard causal overdetermination* – e.g., firing squad cases – where the mechanisms of the overdetermining causes are separate,[106] and b) *constitutive causal overdetermination* (or macro-micro overdetermination) – e.g., the stock market crashes because some investors sell off their stocks in a panic – where overdetermining causes work through the same mechanism (Funkhouser 2002).[107]

The standard type of overdetermination is usually taken to be more problematic because individually sufficient causes bring about an effect through independent mechanisms. Independent mechanisms mean independent causal sources and distinct ways of bringing about the effect. Constitutive overdetermination seems less contentious because overdetermining causes work through the same mechanism – so the feeling of overdetermination is less strong: there is only one causal source and a unique way of bringing about the effect. Considering the distinction from a different angle, constitutive overdetermination is less problematic because it boils down to whether reductionists are right or not: if macro-causes reduce to micro-causes, overdetermination issues disappear. On the other hand, reductionism is not an escape route in cases of standard overdetermination.

The relevant type for group agency realists seems to be the less-disputed constitutive causal overdetermination (macro-micro overdetermination). But group agency realists are not phrasing their view in clear-cut terms. In general, if the accent

[106] The standard type divides into cases of preemptive overdetermination (with temporally ordered causes), concurrent cause overdetermination (simultaneous causes) or asymmetrical overdetermination (simultaneous causes, but having a distinct weight and contributing asymmetrically in generating the outcome).

[107] There is also an iterative overdetermination – when two properties operate through the same mechanism, e.g., a sleeping pill's dormitivity and its chemical property P both cause A's falling asleep (Funkhouser 2002). The distinction between standard and constitutive overdetermination will suffice for the current purposes. For an elaborate taxonomic framework of causal overdetermination types – and actually the most detailed discussion of overdetermination in the literature – see Bernstein (forthcoming-a).

falls on defending the compatibility of group agency realism with a version of individualism, the tendency is to discuss the macro-micro relation and the way groups operate through the individuals constituting them. If on the other hand the accent falls on the autonomy of group agents, the tendency is to think they operate through their own independent devices. This ambiguity stems from realists' fine-tuned position with respect to reductionism, i.e. they want their view to occupy a middle ground and be both individualistic and non-reductive. However, this comes with a price:

a) If the relevant case for group agency realists is constitutive causal overdetermination and group agents and individuals share the same causal mechanism, there seems to be no *independent* sense in which group agents make a difference in the world. In different terms, if groups are causally efficacious through the individuals they are constituted of, there is no independent causal efficacy associated with group agents. Therefore, realists fail to show that group agents play a causal role that is in a sense irreducible to the sum of individual causal contributions $c_1+c_2+\ldots c_n$ of all members – so groups fail to meet the third condition for agency because they lack the capacity to 'actively intervene in the world'. In addition, it may be argued that any case of macro-causation is ultimately a case of micro-causation, and only our causal talk allows the mentioning of entities belonging to various ontological levels or categories. But causal talk liberalism with respect to causal relata does not entail real differences in causal efficacy.

b) At the opposite end of the spectrum, if group agents are autonomous in a certain sense and causally efficacious through their own independent devices, realists will be committed to the standard – and apparently the most problematic – type of overdetermination. Relevant objections here are for instance that we have no reason to overcrowd our ontology and offend Ockham's Razor, that it is mysterious how macro-causes happen to coincide with micro-causes, or simply that we have no reason to believe in some extra entities (on parsimony grounds).

These points show that overdetermination and reductionism are connected issues (at least when considering overdetermination as *prima facie* problematic). Nonetheless, it is difficult to quantify the degree of autonomy realists are hoping for. When being critical about the thin realism of the authorisation theory (the view claiming that group agents are merely authorised by a majoritarian process), it is obvious that group realists

hope for a more robust notion of agency, something more than a reflection of the processes occurring at the level of individuals:

> If a group agent is to display the rationality that agency requires, its attitudes cannot be a majoritarian or other equally simple function of the attitudes of its members. The group agent has to establish and evolve a mind that is not just a majoritarian or similar reflection of its members' minds; in effect, it has to develop a mind of its own. This gives rise to the kind of autonomy that we ascribe to group agents (List and Pettit 2011: 8).

> [W]e must think of group agents as relatively autonomous entities – agents in their own right, as it is often said, groups with minds (List and Pettit 2011: 77-78).

So List and Pettit are clearly after a more robust notion of agency. The phrase 'mind of its own' suggests that groups have intrinsic intentionality and a significant level of autonomy; but compatibility with individualism makes group agents have only derived or extrinsic intentionality. However, the notion of autonomy hinted at is then immediately qualified:

> While the agency achieved by a group supervenes on the contributions of its members – while it is not ontologically autonomous – it is autonomous in another, related sense (…). The autonomy we ascribe to group agents under our approach is epistemological rather than ontological in character and has two aspects, negative and positive. We have seen that to gain knowledge of group agents is to make an important advance in learning about the social world and how to intervene in it (our positive claim). And we have argued that this knowledge is unavailable in practice – even the most idealized practice – on the basis of observing individual agents alone (our negative claim). (List and Pettit 2011: 76)

But a concept of autonomous agency ascribed to groups for methodological reasons – because it would open up the possibility of interacting with and making demands on groups, or because knowledge and understanding of groups would be impossible if we focused only on individuals – is far from the robust notion implied above. In fact, this sense of autonomy does not entail that group agents make a difference in the world *independently*. Yet if this is the sense intended, further argument is needed to show that standard overdetermination is not objectionable.

2.3 Overdetermination Is Not Problematic

On the other hand, why think overdetermination is so bad? Why not say that both group agents and their individual constituents cause particular effects and are morally blameworthy for specific outcomes? A common objection is that overdetermination overcrowds the world with additional causes. Yet an expansive ontology is not necessarily an erroneous ontology – or at least one needs to show that it is erroneous. Another argument could invoke reasons of theoretical simplicity, but if

overdetermination is run together with reductionism, a crowded ontology does not have a higher theoretical price than exclusion (Bernstein forthcoming-a).

An alternative worry is that any action has two causally independent sources coinciding miraculously, or that macro-causation is systematically accompanied by micro-causation. Although no acceptable answer can be given in cases of standard overdetermination, there is no miracle or coincidence that macro-causation is accompanied by micro-causation in cases of constitutive overdetermination, given the necessary truths correlating them (Sider 2003).

Another worry pointed out by Sider (2003) is that we might have no reason to believe in higher level overdeterminating entities. For instance, expressing puzzlement about the need to have higher level regularities or macro-entities in general, Fodor writes:

> [I just do not] see why there should be (how there could be) macrolevel regularities *at all* in a world where, by common consent, macrolevel stabilities have to supervene on a buzzing, blooming confusion of microlevel interactions… So, then, why is there anything except physics? (...) Well, I admit that I don't know why. I don't even know how to think about why. I expect to figure out why there is anything except physics the day before I figure out why there is anything at all, another (and, presumably, related) metaphysical conundrum that I find perplexing. (Fodor 1997: 161).

Yet not seeing a reason for the macro-level regularities or entities is not an argument against the existence of macro-level regularities or entities. As Sider (2003) observes, ordinary beliefs generate some epistemic pressure to postulate macro-entities and, given composition principles, they are in principle possible. Therefore, overdetermination may not be as objectionable as previously envisaged. But in my view, even if that is the case, group agency realism is still not on safe ground. Here are two reasons why:

a) While certain kinds of overdetermination may not be problematic in general, standard causal overdetermination remains objectionable. Fortunately, in the world we live, this is not the most widespread case of overdetermination: we seem to be more interested in various cases of macro-micro overdetermination or in the relation between the mental and the physical, rather than being obsessed with a world in which any event is also caused by overdetermining causes. But unfortunately, as shown above, it is not clear what kind of causal overdetermination is ultimately relevant to group agency realism (or to similar non-reductionist views).

b) However, a more important point is this: even if overdetermination is not objectionable, one still has to explain the causal efficacy of group agents and their relation to the causal efficacy of individual determiners. In different terms, if the problem is not about causal overdetermination, the problem is about how to hold that both group agents *and* individual determiners are causally efficacious. It is not enough to be committed to the claim that both are causally efficacious in virtue of our surface causal talk (*as per* [CAUSAL LIBERALISM]); agency realists have to show *how* they are causally efficacious. It is not enough to answer simply that group agents are causally efficacious trough the individuals they are constituted of, as such an answer would reveal a hidden reductionist manoeuvre and group agents would fail the test of efficacy. I think we should not be surprised with this result. Any theorist would accept a working notion of group agency, but not in the robust sense non-reductive realists are hoping for.

Perhaps the most relevant criticism is that List and Pettit's realism (or any view of that type) plays both sides: it gives the impression of an account marked by courageous 'ontological' bouts – 'group agents are autonomous entities', 'agents in their own right', or 'groups with minds of their own' (List and Pettit 2011: 77-78) – but fuelled by constant qualification and toning-down. Undoubtedly motivated by theoretical fine-tuning, this decisive ambiguity does not help with locating and maintaining a coherent position with respect to causal efficacy. A related objection is that such a position is too finely tuned for a sound definition of related notions such as agency or autonomy: recall that group agency realism occupies a middle ground between authorisation theory and animation theory (Section 1.1, this chapter). What would the definition of agency or autonomy then look like if it is *not* fully independent from individual determiners and *not* authorised by individual members?

2.4 Moral Blameworthiness

Certain aspects of causation are relevant to the way we ascribe moral responsibility to groups. For a group agent to be morally blameworthy, it must satisfy three requirements: it needs to be faced with a morally significant choice (e.g., right vs. wrong), to understand and access the evidence necessary for moral judgment, and to exercise control over their choice (List and Pettit 2011). The causal connection between actions and outcomes is not considered, as moral responsibility is in this context

associated with the ability to make moral judgments and the idea of *control*.[108] If the moral responsibility of collective entities depended on a causation requirement or a test of causal efficacy (as I think they should, at least as far as responsibility for outcomes is concerned), the questions about overdetermination will also reflect on the adequate locus of responsibility ascriptions (groups, individuals, or both). However, leaving that aside, it is important to point out that even the defence of the control condition draws on the notion of multi-level causality. Consider the following individualist argument:

(1) Whatever a group agent does is done by individual agents.
(2) Individuals are in control of anything they do, and so in control of anything they do in acting for a group.
(3) One and the same action cannot be subject both to the control of the group agent and to the control of one or more individuals.
Therefore:
(4) The group agent cannot be in control of what it does; such control always rests exclusively with the individuals who act for the group. (List and Pettit 2011: 160)

The supporter of group agency realism accepts (1) and (2), but rejects (3) by appealing to an analogy. There can be higher-level and lower-level factors that are causally relevant to a certain effect. If a bottle breaks, both the fact that the water in it was boiling *and* the position and momentum of a molecule triggering a break in the surface of the bottle can be said to cause the breaking. In different words, the boiling of the water makes it inevitable that some molecule will trigger the breaking. Analogously, group agents control for the performance of some individuals (and make their action 'inevitable') through the maintaining of procedures, role hierarchy and task division – so both group agents and individuals control the results and are morally responsible.

But the analogy looks deficient. First, the breaking may indeed be described in many ways by citing events, facts or individual entities, but that does not mean that, for example, an event (the boiling) *and* an individual determiner (the molecule *x*) both do the causal work. Levels of description do not imply levels of causal efficacy. Second, there is an equivocation on the notion of control: it is one thing to maintain the procedures or role divisions (group level), and another thing to carry things out (individual level) (recall the distinction between causes and conditions). The general

[108] List and Pettit distinguish their notion of moral responsibility from causal responsibility with a simple example: 'We might hold the dog causally responsible for soiling the carpet, but we would not hold it responsible in the sense we have in mind.' (List and Pettit 2011: 154) – that is, morally responsible. But they fail to mention causation as a prerequisite for moral evaluation: if we find a broken vase and the question is whether to blame the guest or the dog, we would not automatically blame the guest just because she is the 'right kind' of candidate for moral blame and the dog not. We would not treat them differently. On the contrary, we would start from an evaluation of causal sequences and maintain the ascription of responsibility as an open option until a relevant causal connection is established.

intuition is that (more) responsibility is associated with carrying things out rather than preserving an internal structure of hierarchies and task divisions. In addition, it might be argued that it is individuals who designed and maintain these procedures, hierarchies, or role divisions, so ultimate responsibility should be associated with them. Third and related to the last point, the boiling of the water makes it inevitable that some molecule will trigger the breaking, but the maintaining of procedures or role divisions can be done *without* actually carrying things out. On the contrary, individual members may disagree with and act against the groups they are a part of.[109] Fourth, the analogy is top-down and seems to assume group control rather than argue for it – perhaps 'group control' is a mere metaphor for the controlling power of a few significant individual members. And why 'control'? If there is a sense in which groups can also be said to *cause* an outcome, it does not automatically mean they also *control* it. Lastly, the problem of overdetermination could reappear associated with the notion of control. We would then have an issue of overdetermining control, which would simply reiterate the complications of overdetermination but this time related to the ascription of moral blameworthiness.

Group agency realists recognize that the responsibility of individuals as enactors of a group's deeds (rather than as group members or designers of organisational structure) competes with the notion of corporate responsibility and renders it redundant. Yet they believe we should also hold groups responsible anyway, as failures of individual responsibility within groups might go undetected (Pettit 2007). However, corporate responsibility becomes a mere conceptual tool to get access to what ultimately matters – individual responsibility – and as such it hardly justifies the demanding theoretical detour through a defence of group agency. Another possible function of corporate responsibility is regulation of group behaviour: we should hold groups responsible as an exercise in managing or disciplining collective attitudes and actions. However, regulation of group behaviour represents a good reason to attribute responsibility individually. We want individuals to have reasons for regulating their behaviour such as to influence the behaviour of the group (Ludwig 2007) – so the regulation measures would make sense if they were implemented bottom-up rather than top-down. Finally, the sense in which group enactors have more causal relevance than

[109] Alternatively, it could be said that the maintaining of procedures or role divisions raises the probability that some individuals will carry some action out; in contrast, the boiling of the water makes it necessary that some molecule will trigger the breaking.

the designers of organisational structure is unclear. This issue should be reconsidered given the problems surrounding the distinction between causes and conditions.

2.5 Summing Up

There might be disagreement about the irreducibility criteria, however, it is widely held that at least in a certain sense groups can be real and significant unitary agents over and above their individual members. If group agents exist as relatively independent entities and are not mere epiphenomena, their existence is tied to their causal efficacy. Therefore, non-reductive views need a coherent causal story about how group agents impact the world. An argument against group agency realism was that if macro-entities like group agents existed, they would causally overdetermine any effect they may have on the world, since those effects would also be caused by the individuals groups are formed of. If overdetermination is objectionable, either there is no independent sense in which group agents are causally efficacious, or group agency realism is committed to a problematic instance of overdetermination. If overdetermination is not objectionable, one still has to explain the causal efficacy of autonomous group agents and their relation to the causal efficacy of individual determiners. Lastly, for a group agent to be morally blameworthy it needs to satisfy a series of requirements. However, the causal efficacy requirement is not satisfied and the control condition seems based on a deficient analogy. It follows that non-reductive notions of collective agency and responsibility need to provide an adequate causal story about how group agents are efficacious as relatively independent entities. It is fair to say that a) group agents should make an independent causal contribution; if they do not, their existence may be compromised; and b) the independent causal efficacy of group agents will threaten individualism, while causal efficacy that is too dependent on individual members will threaten non-reductive views.

3. A Reply

There is, however, a different way of arguing for the causal autonomy of the systems studied in the special sciences. The first step is to use a view espoused by the British emergentists and some non-reductive physicalists, who claim that since higher-level properties typically have multiple physical realisations, they cannot be identical with the physical properties realising them and therefore possess causal powers independent

from the causal powers of their physical realisers. The second step is to find an account of causation able to establish the autonomy of special-science properties, *viz*. a difference-making counterfactual theory of causation (as used for instance by LePore and Loewer 1987, Horgan 1989, Crane 2001, Raatikainen 2006, *inter alia*). In particular, Menzies and List (2010) appeal to Woodward's (2003) interventionist theory (a version of the counterfactual analysis of causation)[110] and show that under certain conditions causal claims about higher-level properties are *realisation-insensitive*, i.e. they are true regardless of the way they are realised at the lower level. Consider the causal relation between C and E, where C is a higher-level state realised by a lower-level state Ci, but which could have been realised by any other C1, C2, …, Cn lower-level states. The causal relation C → E is realisation-insensitive just in case E is true in some close ~Ci-worlds that are still C worlds, i.e. E remains true if C is differently realised. The third step is to adapt the argument to support the causal autonomy of group agents (higher-level social entities) which is, naturally, a straightforward exercise. For instance, imagine a well-structured group in which multiple distinct configurations of individuals H1, H2, …, Hn correspond to the same higher-level intention I – e.g., Russia's intention to annex Crimea. In the actual world, Russia can be described by a particular complex lower-level configuration H289, which leads to intention I and the performing of annexation action A. However, had its lower-level complex configuration not been H289, it would have been in one or other possible configurations instead – say H467 or H672 – so it would still have had intention I and performed annexation action A. Therefore, the causal relation between I and A is realisation-insensitive and the collective entity as such can be said to make a difference in the world independently of the way its individual members happen to be organised.

This line of reasoning is contentious, however, we can learn some interesting lessons about groups and causation by showing why. For the sake of convenience, I will separate the more metaphysically-oriented observations from the moral aspects entailed by the concept of causal autonomy.

[110] According to the interventionist theory, causal relata C and E are variables and causation relates changes in C-variables to changes in E-variables. In different terms, variable C causes variable E just in case if the value of C were to change as a result of an idealized manipulation (called *intervention*), then the value of E would change as well. On this theory, causes are devices for intervening on or manipulating effects.

3.1 Causal Autonomy and Explanation

The first observation, which I assume to be immediately noticeable, is that causal autonomy is in this context necessarily connected to a *dependence* rather than to a *production* account of causation (Hall 2004). Dependence is the linchpin of counterfactual analysis: a cause C is something that makes a difference to an effect E just in case E *depends* on C, that is, had C not occurred, E would not have occurred. On the other hand, C *produces* E when it helps to generate or bring it about. It is widely controversial whether counterfactual dependence or production are individually sufficient for causation, but this does not deter fans of production views from arguing that difference-making causal autonomy is not enough to establish 'authentic' causal autonomy. To them, genuine causal autonomy is about an independent source and a distinct vehicle of causal efficacy, however, as long as lower-level events do the actual productive work, higher-level events cannot be truly said to be causally autonomous. Therefore, although higher-level properties may prove to be realisation-insensitive, real causal efficacy is still associated with any of the possible lower-level configurations of events.[111]

A second concern is that such a reply would conflate causation with causal explanation: higher-level entities or events may have a different and sometimes more appropriate explanatory role, but not real causal efficacy. When Menzies and List claim that 'the higher-level properties [of systems studied in the special sciences] have causal powers that are independent of those of their more basic physical properties' (Menzies and List 2010: 108), they mean that higher-level properties are described at a different level of specificity than lower-level properties, and that something counts as a cause only if pitched at the 'right' level of specificity.[112] For instance, it would be scientifically inappropriate to causally describe the kinematic behaviour of lenticular galaxies by citing the position and velocity of each star. The specificity principle is similar to Yablo's (1992) proportionality constraint on causes – the idea that causes need to be described with the right amount of detail.[113] However, the specificity

[111] An interesting question is whether it is possible to have realisation-sensitive causal relations even in the presence of multiple realisability. If such examples could be found, multiple realisability would fail to be a sufficient condition for the autonomous causal powers of higher-level events.

[112] Obviously, the interventionist framework is used to ensure the presence of the right kind of counterfactuals.

[113] For instance, when looking for the cause of an injury, being hit by a yellow car conveys too much detail, being hit is not enough, and being hit by a car is just right. What makes for the appropriate specificity depends on pragmatic rules.

principle merely shows that since lower-level explanations of certain phenomena are sometimes inadequate, special sciences enjoy *explanatory* autonomy; it does not show that special sciences enjoy *causal* autonomy.[114]

Indeed, this seems to be the sense in which e.g., List and Spiekermann (2013) understand the reconciliation between methodological individualists (those who reduce social phenomena to individuals and their interactions) and holists (those who take higher-level social entities, events, or properties to be causally significant): although social phenomena are caused by individual-level behaviour, non-reductionist explanations are mandated when these phenomena are 'robust to changes' at the lower-level (where robustness to change is another way of describing realisation-insensitivity). In different terms, if facts about social phenomena supervene on facts about individuals, supervenience individualism does not entail causal-explanatory individualism and it is fully compatible with causal explanatory holism. Although convenient, this view still harbours a tension between dependence and production accounts and for that reason raises further questions: what is the theory of causation actually/ultimately substantiating causal explanation? If different causal explanations correspond to different levels of description and are backed up by distinct causation theories, is there any priority order between them? Are there any principles defining their compatibility? If considerations of causal overdetermination prove to be relevant, would dependence-causation at the higher-level conflict with production-causation at the lower-level?

A third observation is that causal autonomy seems to be secured because groups are complex entities, realisable in countless distinct ways. But once we consider less complex collective entities, the number of potential realisable configurations will reduce, which means that causal autonomy will be a matter of degrees and directly proportional with the complexity of groups. Does this sound right? Imagine that one starts from a simple set of one higher-order property realised by two lower-level properties and continues to add potential realisers (a simple pair of one higher-order property and one lower-level property being excluded by the definition of multiple realisability), at what point does mentioning realisation insensitivity and causal autonomy switch from inappropriate to appropriate?

[114] I am not sure if this would even be possible methodology-wise, as a constraint on relata will have to entail a different, independent way of bringing about an effect.

3.2 Moral Aspects

As a difference-making centred reply, List and Menzies's (2010) defence of causal autonomy should also be consequential for certain moral concepts and views. For instance, such a defence would entail that certain organized or institutional collective entities with authoritative decision procedures (universities, corporations, states) are causally efficacious as group agents and therefore capable of attracting responsibility ascriptions independently of their members (Copp 2006, 2007, Erskine 2003, List and Pettit 2006, Pettit 2003, 2007, 2009, Pettit and Schweikard 2006, Tollefsen 2002a, 2002b, 2003). More clearly, although the members of a group may not satisfy individually the conditions for moral responsibility, the group as a unit may satisfy them and thus achieve a distinctive moral standing.[115] But is difference-making well suited for the work causation does in moral theory? In some situations, yes. For instance, Sartorio (2005) suggests that a conception of causation as difference-making is particularly helpful, e.g. when accounting for responsibility in certain cases of moral luck. Consider the following example:

> Victim is trapped on the tracks. I want Victim to die, and I have reason to believe that the main track is disconnected. So, thinking that the train will derail if it continues on the main track, I flip the switch. As it turns out, however, the main track has never been disconnected. As a result of my flipping the switch, the train turns onto the side track [B], but then the tracks reconverge and the train hits Victim (Sartorio 2005: 92).

The idea could be represented as follows:

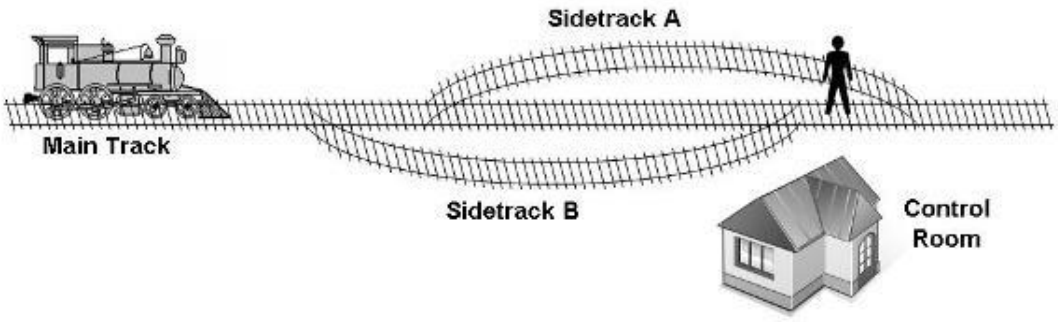


Figure 1. Responsibility and Moral Luck[116]

[115] Naturally, this does not mean that a necessary condition for collective responsibility accounts is that none of the group members has individual responsibility, or that collective responsibility completely excludes individual responsibility. The thesis is that collective agents can be held responsible over and above their members. A particular corollary is that there can be gaps of responsibility between collective agents and their members, that is, cases in which a collective is responsible and none of its members are individually responsible. But lack of individual responsibility is not a necessary condition for collective moral autonomy.

[116] http://www.epjournal.net/blog/2011/04/fun-with-causality-and-morality/

This is a case of moral luck because although I acted wrongly when flipping the switch to side track B, I am not responsible for causing Victim's death (the Victim would have died anyway). Production-based theories of causation, according to which my determining the causal route to the outcome is sufficient for causing the outcome, would hold that my flipping the switch is after all a cause of Victim's death. However, these theories would fail to account for my moral luck, so a difference-making theory would be more helpful in this case.

Yet in other situations, a different intuition takes prominence – the intuition that carrying things out attracts more responsibility than simply belonging to a group that makes a difference. We tend to associate responsibility with those individuals directly involved in bringing about a particular outcome, *despite the fact that* our causal talk is about group-dependence. This tendency can be correlated with the belief in the existence of certain 'zooming strategies' that would reveal the causal efficacy of individuals' reasons and capacities hidden behind higher-order talk. For instance, such a zooming strategy could show how group actions are guided by logical processes taking place in individual heads rather than collective minds. Or it could distinguish between designers of organisational structure and enactors of a group's deeds, between the mastermind and the getaway driver. In such situations, production theories seem to describe more adequately the causal connections relevant to responsibility ascriptions (especially if reinforced by a correspondence relation connecting moral blameworthiness to the part of the outcome one directly produced).

A different observation is that if higher-level social entities were causally autonomous and able to attract responsibility ascriptions independently of their members, moral evaluation would be in a way trapped or left to bounce between levels. The missing premise here is constitution. Add that, and it is again a question of moral luck if group agents or the individuals they are constituted of happen to be morally responsible. Why would individuals be morally responsible for an outcome if group agents would have caused it anyway? And why would responsibility be associated with collective entities if the outcome would have been caused by a particular configuration of individuals anyway? The obvious escaping route is to claim that *both* group agents and individuals are morally responsible, but such a view would have to associate causal

efficacy only with individuals (*as per* Section *3.1*) and therefore sacrifice the causal autonomy initially argued for.

The last remark concerns the way we assess the magnitude of causal contribution and as a consequence, the weight of moral responsibility. Assessing the normative status of group agents, List and Pettit (2011) distinguish between the responsibility of individuals as designers of a corporate entity, as members, and as enactors of a group's plans. Intuitively, their causal contribution to the occurrence of an outcome differs in magnitude, however, there seems to be no unambiguous way of assessing it. There are two basic ways of measuring causal involvement, one in terms of 'more or less' substantial contribution, and another in terms of accurate proportions, say 8%, 28%, 64%. The first kind of measurement is obviously inaccurate and subjective – even if we often have 'a very good, imprecise grasp of size or importance of contribution' (Mumford forthcoming). The second kind is not appropriate for or admitted by higher-level social phenomena. In the '70s, the difficulty of quantifying causal contribution has been used as an argument against the objectivity of social sciences (see for instance Frey 1976 in the Martin-Frey exchange [Martin 1974, 1978a, 1978b, Frey 1974, 1976, 1978a, 1978b, 1978c]) and currently there is no well-developed theory of causation to account for these differences in magnitude. The literature includes a series of attempts at describing degrees of causation (Braham and van Hees [2009], Moore [2009, 2012], Northcott [2013a, 2013b], Halpern and Hitchcock [forthcoming]), but the idea that causation is a scalar notion remains highly controversial. As Moore (2012) observes, it is probably fair to say that scalarity will become clearer once the nature of causation becomes clearer. The lesson to be learned here is that were we able to precisely quantify the degrees of individual causal contribution (and responsibility), a group's causal autonomy would become a redundant notion.

3.3 Summing Up

I have examined a different way of arguing for the causal autonomy of group agents (or, more generally, of the systems studied in the special sciences) and argued that we may learn some interesting lessons about groups and causation by bringing to light its controversial features. In particular, non-reductionist-oriented views will have to:

a) deal with the tension between production and dependence accounts of causation,
b) clarify the distinction between causation and causal explanation, and
c) consider the concept of causal autonomy in relation to various types of groups and their complexity.

As for the moral aspects, non-reductionist-oriented views will need to consider:
a) various issues of moral luck triggered by the idea of constitution and the difference between production and dependence accounts, and
b) whether a potential success in describing degrees of causation will not render group agency realism a redundant notion.

4. Fictionalism

4.1 Motivation

Despite their problems, the views associating agency with groups and describing them as real and robust entities, irreducible to the sets of networked individuals they are constituted of, are well-motivated and valuable. To their credit, such views are logically possible (despite being apparently counterintuitive). In principle, eliminativism entails individualism, but individualism need not entail eliminativism. It is a different question whether individualism may be of a permissive kind, such as to allow irreducible group agents. In addition, such views are supported by the fact that well-structured groups showcase some irreducible features: a) they preserve their identity through changes of membership, b) their reasons for action are not reducible to the reasons of individual members, and c) there is a class of predicates which can only be true of collectives, like 'elected a presidency candidate' or 'passed a constitutional amendment.' These characteristics show that certain groups are often understood in a *non-distributive* sense (transcending the contributions of individual group members) rather than in a *summative* sense (as the aggregation of individual members involved in negotiating a group decision or action). For instance, there are circumstances in which the harmful consequences of an action are widespread and cannot be attributed to specific actions or inactions of particular individuals, but to groups themselves. An oil spill, a war, corruption in a government agency, unreasonable policies, or the unethical or illegal

manipulation of elected officials by interest groups[117] are examples of situations in which action and blameworthiness are associated with collective entities (governments, companies, institutions).

It is important to note that a theory of group agency facilitates social explanation, prediction, and design somewhat in the same way Dennett's (1987) intentional stance[118] helps us understand how certain systems operate. From a normative perspective, a theory of group agency is relevant to issues in political philosophy, economics, jurisprudence, or social policy. Groups are ubiquitous in our society and it is often important that they manifest integrity, reliability and stability, i.e. that they speak with one voice and act with one mind (List and Pettit 2005). If they do, their independent standing is relevant for decision theorists concerned with judgment aggregation and *n*-person prisoner's dilemmas in which groups face difficult choices and have to decide on the right strategy. It is important for applied ethicists working in business (assessing corporate action), medicine (evaluating health care teams) or political theory and IR (evaluating nations and governments). Conceiving states, institutions or international organizations as agents allows us to clarify complexities in policy-making in a world where calls to action come before the identification of the adequate actors. (Erskine 2003). If politics are permissively defined as 'the authoritative allocation of values for a society'[119] and these values are shaped and distributed by the participants in the political process (organizations such as governments, formal institutions, national corporations, interest groups etc.), it follows that group agency represents a key notion in the political domain.

4.2 Two Controversial Assumptions

Amongst the various justifications offered in support of non-reductive views, two assumptions stand out: a) the assumption that groups could meet the conditions set for individual agency, and b) that structural considerations entail group agency. However, both assumptions are controversial.

---

[117] Risser's (2006) examples.

[118] Adopting the intentional stance towards a group (e.g., a government or nation) means to treat it as a rational agent, imagine what beliefs and desires it ought to have given its status and purposes, and predict how it will act in light of its beliefs and desires. After all, as Clark (1994) remarks, most of the international news are broadcasted as a high-stakes soap opera with moody, impulsive and heavily armed players.

[119] Easton (1953: 129).

Assumption 1. I have organised my discussion around the general conditions of agency that any system should meet, i.e. the conditions reflecting a broad consensus amongst psychologists, economists or philosophers of mind: representational states, motivational states, rationality, and causal efficacy. But after further reflection, it might be that the constraints set forth for group agency generate a mere theoretical framework for the ascription of higher-level attitudes, actions, and responsibility, but practically speaking groups are not agents (Corlett 2001). In different terms, while it is clear that collectives *could* satisfy these requirements, it is not obvious that they actually *do*. Take for instance the condition that groups should have representational and motivational states, i.e. that they are intentional agents. Are groups intentional agents? But how could they be, given that they lack minds? They could perhaps be described as *secondary* agents, but their intentionality would be conventional and derived from the intentionality of *primary* agents, i.e. individual members (Copp 1979). They could act for a reason, but it is not clear whether it is their reason or the reason of a powerful individual member or of a faction within the organization. They could act *via* the rules established in an organization, but such rules may describe its *behaviour* without saying anything about its *intentionality*. These points aim to substantiate the idea that if we set forth a number of requirements for non-reductive views, groups could meet them on the theoretical road to agency, but not as a fact.

Assumption 2. We also seem to assume that organizational complexity (the system of rules and norms structuring a collective) reflects or helps to incorporate a group agent in a manner analogous with the way our cognitive architecture reflects a conscious mind. After all, only well-structured conglomerates are said to be agents, and their action is directly determined by their structure. A functioning internal decision structure of rules which coordinate, subordinate and synthesize individual actions was considered enough to grant organizations full-blown agency (e.g., French, 1979, 1984), and much of the literature on legal corporate action follows this idea. Understanding groups as agents owes a great deal to the marriage between *structure* and *function* present in the holistic accounts of social institutions. These accounts relied on the analogy between organizations and complex organisms (Spencer 1860), an analogy which was further applied to governments or nation-states and which generated a conception of groups as embodied structures of roles and related norms. It is important to note however that norms of an organization do not define or fully determine action,

but only *guide* it. More precisely, collective action is not automatically entailed by an excessively formalistic device, a compact cluster of rules and norms complex enough to generate a form of agency. On the contrary, it is better to describe collective action as taking guidance from the norms within an organization. The basis of collective action is in fact more unstable than the collective agency theorist would like to think. To understand this idea it is perhaps useful to consider the distinction between rules in theory (e.g., Rousseau's project of a constitution for Corsica, a legal model on the paper) and rules as functioning inside of an organization, being a part of its social reality. This sociological, anti-formalistic constraint requires us not to understand the organizational norms as antecedently defining individual decision and action, but to understand individuals as taking guidance from these sets of regulations within an organization. As a result, structural properties of collectives do not serve to embody an agent in the sense of fixing its identity and fully determining its actions. Lastly, those still unconvinced should try a thought experiment recommended by Clark (1994: 414). Stop for a moment and grant that certain collectives *are* real and robust agents enjoying a corporate mind. The strategy to block this thought is to imagine yourself shrinking to the point where you could enter this aggregate as if you were entering a mill. From that perspective, one will definitely find unreasonable the idea that a web of interlocking parts could think, act or make decisions on its own.

4.3 A Recommendation

Given its pervasive nature and vital relevance, I would like to offer a brief recommendation concerning our stance towards group agency talk. Facing the fact that what we may call extended notions[120] prevalently inhabit our vocabulary, three choices are open to us:

a) one is to claim that attributions of agency, action and responsibility to organizations are useful fictions, which are strictly speaking false;

b) the second is to claim that they are just euphemisms for individual members' agency, action, and responsibility;

c) the third is to claim that organizations are causally efficacious agents which could consistently be held morally responsible (Tollefsen 2002a, 2002b).

[120] Notions such as collective agency, collective action, collective intentionality, collective responsibility, collective beliefs, desires, attitudes etc.

I have argued that option c) is implausible: it is an error to claim that organizations can be causally efficacious and responsible independently of their members. In response to this error, one alternative is to abandon completely the causal and intentional idiom when referring to organizations. We are therefore left with two alternatives:

Option b): correct the error and maintain the causal and intentional idiom only when referring to minded and rational entities (because only those are the adequate site of agency and intentionality).
Option a): continue as before, but mention a disclaimer which could take the form 'let us pretend that what follows is true, though it is not' (Blackburn 2005).

Against option b), it may be argued that renouncing the intentional idiom would amount to the abandonment of our common sense (for instance it would entail abandoning concepts like 'beliefs', 'intentions', or 'desires' etc.). Neither will it do to restrict the terms of our everyday language such that to refer in intentional terms only to human beings, as we will lose much of our explanatory power. After all, our folk psychology is the source of the error.

It is option a) that I recommend as the most reasonable and appropriate. Our metaphysical and moral stance towards group agency talk should take it as not aiming at the literal truth, but as a useful fiction with nothing than an instrumental purpose. As for the concept of causal autonomy or higher-level causation, it could be best described by appealing to a species of 'as-if' causation usually reserved for omissions and absences – what Dowe calls causation* (2000) and quasi-causation (2001), or what Persson (2002) calls 'fake causation' (assuming that process theories hold). This relation is essentially a counterfactual test about the genuine causation taking place at the individual level, namely:

(1) There is a causal relation between individual members M1, M2, …, Mn and an outcome O, and
(2) Had group G – constituted of members M1, M2, …, Mn non-distributively – not been present, the outcome O would not have occurred.

But what kind of fictions are collective entities? According to Pettit (forthcoming), group agents have been represented as *expressive fictions* by those who take agency ascriptions to be metaphorical; as *pragmatic fictions* by individualists; and as *theoretical fictions* by those who think that positing group agents serves no vital role in social theory. My essentially instrumentalist recommendation is to understand group agents as pragmatic fictions, but at a closer look it seems that Pettit's suggested types of fiction are not mutually exclusive. To a pragmatic fictionalist, collective entities will definitely serve a theoretical role, but not an essential one – so in a sense they also are theoretical fictions. In addition, regardless of what kind of fictions they are, they represent a communicative shorthand, so pragmatic fictions are in a sense expressive fictions. More clearly, pragmatic fictionalists, like any other fictionalists, will treat ascriptions of agency to groups as metaphorical rather than real, and that will make them expressive fictionalists as well. Obviously, Pettit (forthcoming) rejects all versions of fictionalism and defends a strong realist position. Regrettably, none of his arguments addresses the active dimension of group agents to shed some light on their causal efficacy as relatively independent entities.

5. Conclusions

I have argued that recent non-reductive accounts of group talk (and especially group agency realism) focus on the theoretical conditions for group agency and the functional value of organisational design, but fail to provide an adequate causal story of how group agents impact the world as relatively independent entities. I examined the concern that group agents would overdetermine the effects already caused by their constituent individuals and showed that non-reductionists would need a coherent account of higher-level causal efficacy *independently* of whether overdetermination is problematic or not. I then considered a different way of arguing for the causal autonomy of group agents (or, more generally, of the systems studied in the special sciences) and argued that we may learn some interesting lessons about groups and causation by bringing to light its controversial features. Finally, I pointed out the motivation behind non-reductive accounts – in particular their pivotal role to social explanation, prediction and design – and recommended a fictionalist stance towards group agents and group-talk.

The lack of a coherent causal story of how collective entities acquire agential status and make a difference in the world as independent bodies has far-reaching implications for the way we perceive and comprehend the action and responsibility of many voluntary associations, political parties, governmental institutions, commercial corporations, churches, nations, etc. The next chapter takes up a different aspect of causal relations, i.e. their selectivity on the effect side. It explores the difference between effects and by-products, side-effects, long term effects, and provides further insight into the way effect selection is relevant to determining the appropriate scope of responsibility for consequences.

IV. Effect Selection

'Selection (…) applies to just one side of the causal relation: the *c*-side. Though perhaps there is also selection on the *e*-side, between *the effect* and *its byproducts* (or *side effects*). Still, it seems that the effect/byproduct distinction is not quite parallel to the cause/condition distinction since the effect/byproduct distinction seems to have more to do with agential intentions and less to do with the contrastive structure of causal inquiry. I do not have an explanation for this asymmetry'. (Schaffer 2005: 352).

Introduction

Causes have infinitely many effects. There are at least three reasons for this. First, effects are often accompanied by countless complex by-products,[121] which can themselves be conceived as effects. Second, if transitivity[122] holds, it is possible in principle to trace a causal chain indefinitely into the future. Third, consequences may be conceptualised at different levels of specificity. For instance, two tectonic plates colliding and causing an earthquake generate multiple adjacent effects we do not normally mention, e.g. a propagation of elastic waves, a minute remodelling of the ocean's floor, or a modification in the Earth's tectonic configuration. Complexity increases if we follow the causal chain forward. There will be intimately related effects such as seismic sea waves, structural damage or mass wasting events, but also long-term economic or psychological consequences. Finally, ramifications may be depicted in more or less detail, e.g., the cultural effects of an earthquake may be described without mentioning every lost archive or work of art. In theory, every event has indefinitely many consequences, but we choose to cite a few or usually just one in singular causal statements. This suggests that a mechanism of selection operates on the effect side of causal relations: just as causal selection promotes some factors to the status of causes and demotes other factors to the status of mere conditions, effect selection elevates some factors to the status of effects and relegates other factors to the status of by-products, side-effects, after-effects, etc. The analogy is not perfect, however, we may intuitively conceive of the 'real' Effect as the whole of the consequents and take any causal statement to be incomplete unless we somehow introduce all of them.[123] If causes have infinitely many effects, but only one or a few

[121] Throughout the chapter, I will be mainly using the concept of by-product as a generic term for what follows or derives from effects as such. For a brief taxonomy of effects, see Section 2.5.

[122] A relation *R* is transitive iff if *x* has *R* to *y* and *y* has *R* to *z*, then *x* has *R* to *z*, that is, iff ∀(*x*, *y*, *z*) ((*Rxy*&*Rxz*) ⊃ *Rxz*).

[123] For instance, G. E. Moore notes: 'One natural way, and perhaps the most natural way, of understanding the expression "the total consequences of the action, A," is one in which among the consequences of A nothing is included but what is the case *subsequently* to the occurrence of A, so that the "total consequences of A" means everything which is the case *subsequently* to A's occurrence, which

are selectively highlighted and mentioned in causal claims, what determines their selection from the complete set of consequents? The main goal of the chapter is to explore the problem of effect selection and assess its relevance to determining the appropriate extent of responsibility for consequences. I argue that the difficulties in the way of a clear answer to the puzzle of effect selection motivate an epistemic argument against consequentialism.

Overview: Section 1 introduces effect selection, surveys several attempts to formulate selection criteria, and points out some of their problems. Section 2 examines if two popular ways of accounting for causal selection apply to the effect side. A prospective approach to effect selection needs to strike a balance between two unappealing extremes: (a) the claim that effects and by-products are metaphysically distinct, and (b) the claim that there is no sense in which effects and by-products are objectively different and selection is always governed by context-dependent pragmatics. I argue that despite the strong sense that effects and by-products are essentially different, the criteria governing their differentiation are neither clear nor predictable. Section 3 considers the relevance of effect selection to determining the appropriate scope of responsibility for consequences and develops an epistemic objection to consequentialism. Section 4 concludes.

1. Selection x 2

1.1 Selection in the Opposite Direction

The problem of causal selection emerges from several observations articulated by Mill (1843/1947: III, V, 2/3):

(i) The causal course of nature can be seen as a 'web composed of separate fibers', 'a collective order made up of particular sequences, obtaining invariably among the separate parts' (Mill 1843/1947: 213).[124]

---

is also such that it would not have been the case if A had not occurred.' (Moore 1942: 559, emphasis in original).

[124] Mill uses a representation of causal structure as a complex ordered network, narrowing back down to the beginning of the universe and branching openly towards the future. The idea of a network derives from modelling the generation of effects after paradigmatic mechanical phenomena in which the transfer of motion is regulated by the principle of composition of forces. Similarly, effects are the result of particular combinations of two or more causes which add up by a sort of vector algebra – what Mill (1843/1947, VI) calls the principle of Composition of Causes. The resulting network model is a standard metaphysical representation of the most fundamental level of reality, *viz.* a seamless structure of causal interactions that may extend *ad infinitum*, is intricate and dense, and may take different interpretations depending on the ontology of the nodes (events, states of affairs, facts, tropes etc.) and the character of relations (deterministic or probabilistic).

(ii) There seems to be an 'invariable order of succession' between certain events (that is, certain events constantly succeed other events), however, these invariable sequences seldom (if ever) hold between a consequent and a single antecedent (Mill 1843/1947: 213-214).
(iii) The number of conditions causally determining an event exceeds those mentioned in its causal explanation (from [i] and [ii]).
(iv) Despite (iii), we commonly single out an antecedent, the Cause, and call the others mere Conditions.

For instance, a person may eat a particular dish and die as a consequence, but there is no 'invariable connection' between these events. This indicates that if death occurs more conditions need to be mentioned in its causal explanation – 'a particular bodily constitution, a particular state of present health, and perhaps even a certain state of the atmosphere' (Mill 1843/1947: 214). In fact, if reality is represented as a directed, uniform network of causal interactions, the 'real' Cause of death is the complex sum of *all* determinants in the past light cone of the target event. The challenge raised by the classical problem of causal selection is to explain the priority typically given to a condition or set of conditions (*the Cause*) over other determinants (conditions or *background factors*). To Mill, this differential treatment was entirely arbitrary:

> The real Cause is the whole of these antecedents; and we have, philosophically speaking, no right to give the name of cause to one of them exclusively of the others. (…) All the conditions were equally indispensable to the production of the consequent; and the statement of the cause is incomplete, unless in some shape or other we introduce them all. (…) Nothing can better show the absence of any scientific ground for the distinction between the cause of a phenomenon and its conditions, than the capricious manner in which we select from among the conditions that which we choose to denominate the cause. (Mill 1843/1947: 214-215).[125]

Yet selection on the cause side is only part of the story, as a similar thesis seems to hold in the opposite direction. We commonly single out a consequence under the denomination of Effect and call the others merely by-products. However, a parallel Millian stance would visualise the 'real' Effect as the whole set of consequents and consider any causal statement incomplete unless we somehow introduced all of them. Following the analogy, we have no right – philosophically speaking – to give the name

[125] The argument from arbitrariness resurfaces in the recent literature. For instance, Lewis (1973) aims for an analysis of causation unperturbed by principles of 'invidious discrimination'; Hall (2004) distinguishes between a transitive, egalitarian sense of cause and a restrictive, salience-oriented sense; and Beebee (2004b) defends the Millian network model of causation.

of Effect to one of the consequents exclusively of the others. Therefore, the puzzle of effect selection is to explain the priority typically given to a consequence or set of consequences (*the Effect*) over other results (*by-products*). Is there a metaphysical basis for the distinction between effects and by-products, or is it just a discretionary and indiscriminate matter?

To frame the ensuing discussion of effect selection, it is useful to call attention to two basic distinctions. First, as pointed out by Hart and Honoré (1985) or Hesslow (1988), analyses of causation rest on the assumption that questions about *connection* can be treated independently from questions about *selection*. The connection problem is that of understanding the nature of the causal relation, the process or structure of dependence by which conditions determine effects in general. The selection problem is that of weighing the relative importance of relata (particular causes versus background conditions and specific effects versus by-products). Yet this assumption is not obvious and some philosophers caution, if only briefly, that selection of relata is as much part of the meaning of causation as connection is.[126]

A second distinction is that between *horizontal* and *vertical* selection. Horizontal selection is *prioritisation* according to a certain principle. For instance, the effect of antihistamines is the relief of symptoms associated with allergic disorders, a consequence prioritised over side-effects such as drowsiness, dizziness, or blurred vision.[127] Vertical selection is *description at a certain level*, and may involve different procedures such as extraction, abbreviation, sampling, abstraction (in one sense), or zooming in and attention aiming (in the opposite sense). For instance, alleviating the effects of loratadine may be described without mentioning how they physically block H1 receptors, stopping histamine from reaching its target. Although both types of selection operate simultaneously and the relegation of a factor to the status of by-product depends to a certain extent on the level of description, the real problem of selection is that of justifying prioritisation or differential treatment on the horizontal dimension. The reason is simply that we are more familiar with a certain variation in

[126] For example, Hart and Honoré remark: 'The contrast of cause with mere conditions is an inseparable feature of all causal thinking, and constitutes as much the meaning of causal expressions as the implicit reference to generalizations does'. (Hart and Honoré 1985: 12). Also, when discussing negative causation, Lewis remarks: '[N]o theory of the causal relation (…) can be the whole story of causation. (…) Any relation needs relata, whether it is intrinsic or not. So the problem of missing relata hits any relational analysis of causation.' (Lewis 2004: 282).

[127] Although first generation sedating antihistamines may be specifically used for their sedative effects (the so-called 'off-label use').

the description of causal facts, perhaps following considerations of conciseness or different explanatory interests, but less familiar with the reasons for prioritisation (e.g., when disagreeing on causal selection). My focus will therefore be on horizontal selection or the prioritisation of effects over by-products.

1.2 Principles of Effect Selection

What could determine the selection of certain effects from the full set of consequents? Several broad criteria spring to mind:

a) Interest. An intuition favoured by most people is that effect selection is governed by explanatory-related interests. Examining the 1755 Lisbon earthquake, seismologists might focus on effects such as the tidal waves sweeping the Portuguese coast, whereas economic historians could be interested in the number of deaths, the percentage of structural damage, the damaged fleet, or the lost works of art. One may also mention the impact on society and philosophy: to Voltaire (1911), the 1755 earthquake provided a counterexample to Leibniz's idea that our world is the best of all possible worlds;[128] to Rousseau (1967), it offered an argument against cities and for a more natural way of

[128]In his 'Poem on the Lisbon Disaster or an Examination of the Axiom "All is Well,"' Voltaire vehemently criticised Leibniz's ideas:
'Oh, miserable mortals! Oh wretched earth!
Oh, dreadful assembly of all mankind!
Eternal sermon of useless sufferings!
Deluded philosophers who cry, "All is well,"
Hasten, contemplate these frightful ruins
This wreck, these shreds, these wretched ashes of the dead;
These women and children heaped on one another,
These scattered members under broken marble;
One-hundred thousand unfortunates devoured by the earth
Who, bleeding, lacerated, and still alive,
Buried under their roofs without aid in their anguish,
End their sad days!
In answer to the half-formed cries of their dying voices,
At the frightful sight of their smoking ashes,
Will you say: "This is result of eternal laws
Directing the acts of a free and good God!"
Will you say, in seeing this mass of victims:
"God is revenged, their death is the price for their crimes?"
What crime, what error did these children,
Crushed and bloody on their mothers' breasts, commit?
Did fallen Lisbon deeper drink of vice
Than London, Paris, or sunlit Madrid?
In these men dance; at Lisbon yawns the abyss.
Tranquil spectators of your brothers' wreck,
Unmoved by this repellent dance of death,
Who calmly seek the reason of such storms,
Let them but lash your own security;
Your tears will mingle freely with the flood.' (Voltaire 1911: 10-11).

life.[129] And it certainly influenced the 18th century Portuguese politics or the development of seismology and scientific geography. Each person singles out the effects that interest her most. Sometimes, interest guides the interpretation of a phenomenon with multiple effects by providing emphasis – for instance, surveillance in areas that may need monitoring might be interpreted as increasing public safety or as causing loss of privacy and individual freedom. Interest appears to be an appealing selection criterion mainly because it conveys the flexibility of choosing any event as an explanatory target. However, if interest reigns over effect selection, there seems to be no principled problem of prioritisation of effects over by-products, but only a question of working out which interests are guiding discussion.

b) Known and unknown effects. Another reason why we do not enumerate all additional effects deriving from a cause may be that some of them are already known or understood without being communicated. For instance, if an earthquake causes a tidal wave, there is a certain degree of awareness regarding the risks or dangers normally brought about by tidal waves. At the opposite end of the spectrum, a large number of consequences of particular events are unknown or deemed too remote to matter, therefore some effect selection is justified by appealing to relevance-based reasons or denouncing our cognitive limitations. A similar knowledge-based criterion links effect selection to what is taken to be prominent, unusual or unexpected. The rationale here is that we do not usually remark and describe something as an effect (an event or state of affairs in need of explanation) if things behave normally.

c) Proximate effects. Mill (1843/1947: III, v, 3) remarked that we often 'dignify with the name of cause' a condition (or cluster of conditions) immediately preceding an effect or which 'come last into existence'. This selection criterion benefits from support in everyday language, *viz*. talk about deep and surface causes or about precipitating causes. Along the same lines, we frequently define effects as those events which come

[129] Shocked by Voltaire's poem, Rousseau replied in a letter (in August 1756): 'You would have preferred that this earthquake had taken place deep in a desert rather than at Lisbon. Is it possible to doubt that they do not occur in deserts? But we do not speak of those because they cause no harm to the Gentlemen Who Live in Cities, the only people we take into consideration. These earthquakes scarcely harm even the animals and the savages who sparsely populate these remote regions and who do not fear falling roofs or collapsing houses. But what is the significance of such a privilege? Does this really mean that the order of the natural world should be changed to conform to our caprices, that nature must be subject to our laws, and that in order to prevent her from causing an earthquake in any particular place all we need do is build a city there?' (Rousseau 1967: 37).

first into existence (given a cause C) and we encounter similar linguistic support in the distinction we draw between conventional effects and long-term effects, after-effects, side-effects, etc. In both cases, it is important to note that describing an effect as 'long-term' or a cause 'precipitating' suggests an awareness of multiple effects or causal factors, respectively. Several potential tests for the proximate effect criterion may include its immediately foreseeable character or the degree to which it follows directly from the assumed cause.

d) Intention. According to manipulability theories of causation, causes are devices potentially exploitable for manipulation or control: X causes an effect Y if manipulating or changing X would change Y (Gasking 1955, Collingwood 1940, von Wright 1971, Menzies and Price 1993, Woodward 2003). However, it is often the case that, beside the changes intended in Y by the manipulation of X, Y undergoes changes not intended when X is intervened upon. As mentioned above, the administration of antihistamines (X) may be intended for the symptomatic relief of allergies (Y), but some have side effects such as sedation or psychomotor impairment. Therefore some effect selection depends on whether an effect or cluster of effects is intended or not. Occasionally, some drugs are prescribed or procedures are followed precisely for their side-effects; in these cases, side-effects are substituted for the desired consequences. Finally, in another important category of cases, relevant intended effects may come apart from foreseen consequences (which raises critical questions about the permissibility of intervening on X). For instance, the use of anabolic steroids boosts sports performance, while at the same time induces well-known adverse effects ranging from various neuropsychiatric symptoms to significant liver damage.

These are only a few broad suggestions usually invoked to account for effect selection (the list is in no sense inclusive and the criteria may be integrated). It is immediately noticeable that the idea of selection criteria is incompatible with the widespread view that a differential treatment of effects is *arbitrary* (cf. fn. 125). At a closer look, what apparently justifies the search for criteria is the uniformity of our selections, their conventional similarity, or their 'too predictable to be without a basis' character.[130]

[130] In support of predictability, Hart and Honoré note that '[i]n most cases where a fire has broken out the lawyer, the historian, and the plain man would refuse to say that the cause of the fire was the presence of oxygen, though no fire would have occurred without it: they would reserve the title of cause for

Most effect anticipations seem to be based on a naïve regularity view – we usually know what prominent consequences routinely follow from a significant number of events or acts – and as predictions get more complex our theories get more elaborate. On this reading, identifying a rationale for the predictability of our selections *just is* establishing selection criteria, i.e. arriving at principles such as those recommended above. In addition, it may appear that interests represent a more fundamental category, encompassing all the others, given that what we intend, what is proximate, and what is normal are all likely to be objects of our interests. In different terms, the effects defined by other selection criteria seem analysable in terms of interests, but not vice versa. However, care is needed in distinguishing the idea of having a particular, explanation-related interest (such as geological vs. historical or cultural in the case of an earthquake) from the idea of having an interest in a specific selection criterion (e.g., proximate vs. intended effects).

Although initially plausible, these principles are nonetheless problematic (Hesslow 1988). For example, how can we tell a *correct* criterion from a deficient one? Since there is some truth in each of them, it is tempting to go pluralist and hold that different selection principles apply in different contexts. However, this would create the additional problem of selecting an appropriate criterion for each situation. Furthermore, formulating the central issue as one of *selection* presupposes awareness of most or all potential effects an event may trigger – otherwise the notion of selection would not make much sense. This is particularly important to those who suggest that selection is mainly governed by interests, as it would not make much sense if these were the only effects we knew about. But are we aware of the downstream consequences following from different events and acts? Very often, effect selection goes off unhindered and blissfully unaware of any extra consequences. And even if we were aware, awareness is a complex thing – one may be aware of causal consequences, but less aware of logical consequences; aware of proximate consequences, but unaware of remote consequences, fully aware, or merely peripherally aware, and so on. In fact, it is not very clear what we know about the effects which are not selected. If science revealed with accuracy all consequents of a certain phenomenon, would we still

something of the order of a short-circuit, the dropping of a lighted cigarette, or lightning… In making this distinction it is plain that our choice, though responsive to the varying context of the particular occasions, is not arbitrary or haphazard.' (Hart and Honoré 1985: 11).

continue to distinguish an effect or cluster of effects from its by-products?[131] Lastly, it is difficult to point out a *rationale* for choosing a principle of effect selection. Different criteria seem backed up by different rationales. The interest criterion is motivated by the explanatory function of causal knowledge,[132] known and proximate effects criteria are motivated by pragmatic reasons and informational economy, and the intention criterion is prompted by a concern with manipulation, control, and prevention. It is again appealing to assume a pluralist stance, but in many cases effect selection is motivated by more than one principle – for instance, we may target certain proximate effects because of relevance- or information economy-related reasons and because we have a specific interest in mind.

The difficulties affecting principles of selection suggest two lessons for a prospective approach:

(1) Effect selection is problematic along several dimensions: it requires an explanation of the distinct *nature* of effects and by-products and what substantiates the divide between them; it demands offering a *selection axiom* (or different selection rules) to explain why mentioning only one or several effects rather than the entire set of consequents is enough in a recurrent manner; finally, it requires presenting a *rationale* for the selection principles argued for. A well-developed solution of the effect selection puzzle would need to provide an answer to these matters.

(2) An account of effect selection would have to note that it is implausible to understand effects and by-products as belonging to completely distinct ontological categories (given the network model suggested by Mill), and equally implausible to argue that there is no real, objective difference between them and arbitrary interests always dictate (given the scientists' concrete efforts to determine effects and isolate them from by-products – e.g., in the case of most prescription drugs). The ideal approach would therefore need to find a way between these extremes, offer an improved explanation of the phenomenon, and overcome the difficulties discussed above.

[131] I do not have a clear answer to this question. It seems that we continue to make this distinction with respect to causal selection. It might be that we would maintain it in some contexts for effect selection (e.g., in criminal trials).

[132] More clearly, (i) causal information about an effect substantiates the explanation of that effect, and (ii) we might have different interests in different causal information (e.g., in the geology vs. cultural history of an earthquake).

2. Approaching Effect Selection

The goal of Section 1 was to introduce the distinction between effects and by-products and discuss several principles of effect selection. In this section, I examine if two ways of approaching causal selection apply to the effect side: *relativisation* (Hesslow 1988, Cheng and Novick 1991, Schaffer 2005, Waters 2007) and *explanatory economy* (Franklin-Hall forthcoming-a). It is a natural methodological move to import strategies relevant to causal selection and attempt to build solutions on the analogy while regularly controlling for asymmetries. According to a version of relativisation, causes and conditions are in a sense ontologically distinct, which triggers the question whether the same deep difference holds for effects and by-products. According to explanatory economy, the distinction between causes and conditions is guided by principles of explanation, which makes one wonder whether a similar surface difference holds for effects and by-products. I argue that both strategies are problematic and, while there is a strong sense that effects and by-products are different, the criteria governing their differentiation are neither clear, nor stable.

2.1 Relativisation

One way to think of effects as distinct from by-products is to argue for their prominence relative to a given reference frame (cf. Chapter I). Adapting elements from classic analyses of causal selection (Anderson 1938, Mackie 1955, 1965, Hesslow 1988) – a *field*[133] and a *relevance operator* – the idea is to construe effects as deviations or differences within a circumscribed region and then argue that effect selection is determined by relevance-based considerations.

Let me explain. Ideally, a superhuman intellect would be capable to grasp the whole set of networked consequents deriving from a particular cause or cluster of causes; however, given our cognitive limitations, achieving specificity requires the introduction of a certain reference frame (a sort of restriction on the network model that enables us to deal with particular causal claims without including the whole prior or future states of the universe). The idea of a reference frame has taken different descriptions in the literature – causal field (Anderson 1938, Mackie 1965), contrast

[133]The concept of *causal field* was first introduced by John Anderson (1938) to deal with the Millian 'capriciousness' argument about causal selection, but it became more familiar with J. L. Mackie's treatment of the difference between causes and conditions (Mackie 1955, 1965, 1980).

space (Garfinkel 1981), focal set (Cheng and Novick 1991), population (Waters 2007), framework (Strevens 2008, 2013)[134] – and defines a state of affairs that is presumed or held fixed when causes and effects are specified. Causal claims of the form 'C causes E' are thus elliptical and need to be expanded as 'C causes E relative to F' to include an argument place for the reference frame. For instance, a causal claim like 'The Azores-Gibraltar Transform Fault's complex tectonic behaviour caused the 1775 Lisbon earthquake' makes sense against a normal, unperturbed course of events or a general state of affairs S in which earthquakes do not occur. In different terms, it is not that an event X causes an event Y – full stop, but that an event X causes an event Y *given S*. Once a reference frame has been stipulated, an effect is specified as a difference within that demarcated region, e.g. the occurrence vs. the non-occurrence of an earthquake. To illustrate more complex cases, consider two examples of seriatim and parallel effects:

[Seriatim Effects] In *Petition of Kinsman Transit Co*., 2d. Cir. 1964, 338 F, 2d 708, 'the operator of a 500 tonnes vessel failed to properly moor its ship along the Buffalo River. Ice forced the vessel from its mooring, into another vessel. The two ships were then forced against a bridge by the moving water and ice. The bridge failed and the resulting wreckage of the bridge and drifting vessels created a dam, blocking the river. The river backed up, flooding a manufacturing area of the town.'[135] To this chain of results one may add more remote consequences such as court trials, psychological effects on the people in the affected area, etc.

[Parallel Effects] The deliberate administration of a potentially unsafe dose of morphine causes both a relief from unbearable pain (a good effect) and the death of the person (a bad effect).[136]

---

[134] Other relevant terms are 'context' 'domain of interest', 'sampling space' etc.

[135] Example detailed at www.legal.com, http://www.legal.com/law-students/50-perspective-on-causation?start=2

[136] For simplicity reasons, also assume that in both cases effects follow from one cause:
The failure to properly moor the boat → the drifting of the boat → the crash into another boat → the crash of both boats into the bridge → the collapse of the bridge → the blocking of the river → flooding the adjacent land → … → court trials → … → economical and psychological effects → … → a doctrinal change in the American Law → … → etc.
Administration of morphine → relief from pain
→ death

In both cases, effects are understood as differences in particular fields: for instance, the drifting of the boat is a specification in the field of things that may drift and it is conceived as a deviation from a non-drifting state. In different terms, it is a difference between the moment the drifting occurs and the times in which drifting does not occur. Analogously, the pain relief is a specification in a field of things that may be relieved from pain (human beings in this case) and it is conceived as a deviation from a painful state. A relevance operator guides the choice of reference frames. As soon as a reference frame is selected, the attention is set on a certain effect-specified-as-a-difference, while the others become relegated to the status of by-products, becoming absorbed by the field. Therefore, the answer offered by relativisation is that all effects of a given cause are equally salient but we apply a set of relevance-based conditions for selecting a certain field. Effect selection is arbitrary as long as the context is not fixed, however, once a context is chosen, one can expect agreement on effects.[137]

2.2 Are Effects Ontologically Distinct from By-Products?

A particular version of relativisation focuses on the causal reasoning of biologists accounting for the role of DNA in the development of different phenotypic traits and sees causes and conditions as being ontologically distinct. Waters (2007) notices that classical geneticists' causal thinking in experimental contexts naturally appeals to an ontological distinction between the general or *potential* causes of a phenotypic trait (for

[137] From a formal point of view there is no reason to say that a cause brings about an effect rather than another, a proximate rather than a remote effect, a known rather than an unknown event, an intended consequence rather than a side-effect.

C → E1→ E2→ … → En

→ E1
C → E2
…
→ En

Here is where the relevance operator intervenes: if we are interested in a difference between E1 and ¬E1 given the presence of C (with '¬' read as non-occurrence), E2 and E3 become irrelevant. Alternatively, if we are interested in a difference between ¬E2 and E2 given the presence of C, E1 and E3 become irrelevant/part of the field.

→ E1
C → ~~E2~~
→ ~~E3~~

→ ~~E1~~
C → ¬E2
→ ~~E3~~

instance, a trait like the eye colour of fruit flies) and the particular or *actual* difference-makers causing a specific difference in the trait (for instance, the difference between red eyes and purple eyes in fruit flies). According to the classical geneticists, if the question is why a number of fruit flies have red eyes, there will undoubtedly be many factors contributing to the redness, both genetic and environmental. However, it is only by embedding the red-eyed fly in an extended, mixed population of flies with different features – some with red eyes and the same genotype (+, pr) and some with purple eyes and a different genotype (pr, pr) that we are able to distinguish between potential causal determinants and actual difference-makers, i.e. their having a specific version of a gene in that population (an a+ allele). To see precisely why some fruit flies have red eyes, we need to compare them with the fruit flies with purple eyes, within a determined population, so we could identify what actually determines the differences at the genetic level. To put it differently, once we define a target effect as an actual difference within a chosen experimental population (or once we locate a difference of a property relative to a reference frame), we are able to distinguish between actual difference-making causes and mere potential, background determinants. Waters understands the distinction between potential causes and actual difference makers as an *ontological* one, in the sense that actual difference makers introduce a genuine biological difference in a phenotypic trait.

Would the same reasoning apply to the distinction between effects and by-products? Apparently not. Consider what it would mean to say that effects may be described with a similar difference between actual and potential changes. While some changes look like mere possibilities, by-products are, just as much as effects, actual changes. If I strike a match and it lights, the presence of a flame-shaped reflection on my retina is a by-product of the match lighting, but just as much an actual change. Similarly, any further variation in phenotypic traits would be treated by geneticists as an effect of the relevant genotypic configuration. In different terms, any minute change would count as actual (or particular) and none as potential (or general). This suggests that something is deemed an actual change once an experimental reference frame is chosen, otherwise it receives a secondary status and is described as a mere by-product. Therefore something counts as an effect or a by-product depending on what reference frame is chosen. Apparently, then, the key to understanding the difference between effects and by-products is in tracking how reference frames shift. Two issues spring to mind:

a) The reference frame variation in cases of selection of causes vs. conditions is guided by the retrospective explanatory inquiry. For instance, the engineer may explain the accident as being caused by the deteriorated road surface, while the detective may explain the accident as being caused by the drunk driver. The reference frames shift because the retrospective explanatory inquiry into causes vs. conditions shifts. However, there seems to be no comparable forward-looking, prospective inquiry into effects vs. by-products (although one could say that e.g., testing in a population is aimed at securing the recurrence of a desired effect over accompanying and usually detrimental consequences, that is, testing secures a recurrent distinction between effects and by-products). If a retrospective look at the relevant causal history of a particular event is motivated by explanatory aims, what exactly motivates the prospective look at the relevant effects and the way they differ from the rest of consequences? If the emphasis provided by the explanatory inquiry shows why certain causal determinants are left in the background, what justifies the corresponding situation on the effect side?

b) Reference frame variation seems guided by the way we understand the opposite relatum. As Waters (2007: 566) remarks, for a difference maker to exist, there must be a difference, *viz*. a difference amongst entities or in the same entity at different times. The idea is that we need to describe the effect as an actual difference in a population (e.g., a difference in the eye colour of fruit flies) in order to identify actual difference makers amongst potential difference makers (e.g., the purple gene):

> It makes no sense to identify something as the actual difference maker without identifying either a population with at least two entities that actually differ with respect to the effect variable Y or a population of one or more entities that exhibit different Y values at different actual times. It is worth stressing that the actual effect is not a single property in a single token, it is a difference of a property in a population. (Waters 2007: 567).

If correct, however, this observation renders Waters' resolution of the causal selection puzzle dependent on a previous understanding of effect selection (we need to construe the effect as a difference of a property in a population or with respect to a reference frame). It may also imply that effect selection is similarly dependent on a particular way of individuating causes (something is an effect rather than a by-product if it follows from a well-defined cause). If a difference on the cause side depends on the way we describe a difference on the effect side, it might be that a difference on the effect side depends on the way we describe a difference on the cause side. This observation also

ties in with a different issue: is the distinction between potential and actual difference-making causes really an *ontological* one? Waters admits that the selection of difference-making causes depends on the prior determination of a relevant population, but thinks it is an error to infer from the fact that effect selection is interest-relative to the idea that what counts as a real difference maker is also interest-relative:

> The error (…) is to infer from the fact that the selection of an effect in an epistemic context involves pragmatics to the mistaken idea that what counts as the cause of the effect must also depend on pragmatics. My point is that once the effect is fully specified as an actual difference in a real population, the issue of which causes are the actual difference makers is an ontological one. (Waters 2007: 570).

In different terms, we must not take the interest-driven specification of effects to affect the identification of genuine causes. Nonetheless, it is unclear why effect specification can be entirely free and interest-driven, while the distinction between actual and potential difference makers retains an ontological character. After all, some may argue that causal selection is also guided by interest – not directly, as Mill or Lewis argued, but indirectly, given the interest-driven specification of effects. Moreover, readers may perceive the difference between actual and potential determinants as a contrast between factors that change (causes) and factors that do not (conditions), but this does not seem to mark a very deep ontological divide, *viz*. a difference in the nature of determinants. In addition, all causes seem actual in the classic Millian sense that all actual determinants causally depend on potential determinants to bring about an effect. To avoid this interpretation, Waters frames the discussion in terms of variables and proposes that a causal variable plays an actual differentiating role in the sense that actual variations in its value explain the actual variations in the values of the effect variable. Other causal variables cannot play this role as their values are invariable across the actual population. However, a) talk of differentiating roles of variables emphasizes similarity in the nature of determinants rather than ontological difference, and b) talk of differences in values being correlated with and explaining other differences in values seems to relocate difference-making from causation to causal explanation.[138]

Lastly, Waters' distinction between actual causes and background conditions should also work for effects vs. by-products. Since similar dependence relations hold on both sides in virtue of the network model, it is implausible to admit an ontological

[138] Cf. Chapter I for the distinction between causation and causal explanation (Section 2.2) and the general problems of causal modelling (Sections 3.2-3.5).

distinction on the cause side and one of a different nature on the effect side. But things are not that simple. Translated into the language of changes, the ontological difference on the cause side emphasizes factors that change over factors that do not; however, on the effect side the relevant difference between effects and by-products is between factors that change and further changes deriving from them (or distinct changes deriving directly from the considered causes).[139] Were we to preserve an ontological distinction of Waters' type on the effect side, it should be one between effects/by-products and other factors remaining unchanged (a sort of conditions for effects) rather than one between changes. It is therefore unlikely that effects and by-products are ontologically distinct.

In general, relativisation provides a persuasive way to think of effects as distinct from by-products by showing their prominence relative to a given reference frame, but as a strategy it brings up the same problems encountered in Chapter I. First, effect selection seems to go off unconstrained by any reference frame. For instance, when we claim that lack of rain causes the great wildebeest migration in the Serengeti plains (rather than other environmental consequences), effect selection is entirely effortless and no mention to populations, contrast spaces or focal sets is made. In reply, it may be countered that reference frames are implicit in the effect specification, but it is not clear why we select a reference frame rather than another. Second, relativisation should account for the predictability of effect selection patterns – why effects are consistently prioritised over by-products – in terms of recurrent reference frames. According to supporters of relativisation theories, reference frame stipulation follows from the explanatory inquiry in cases of causal selection, but recall that no corresponding forward-looking inquiry into effects vs. by-products is available. To avoid a lapse into arbitrariness, it would have been perfect to have a contrastivist account of effect selection, but if such a parallel solution fails, then no clear principle marking an objective difference between effects and by-products can be stated.

[139] Certain care is is needed here, as not all by-products derive from what we call the effect. Some derive directly from the cause(s). For example, a prescription drug's side effects (say, drowsiness and nausea) need not derive from its intended effect (pain relief), but will rather be directly caused by the prescription drug itself.

2.3 Explanatory Economy

A different way of approaching effect selection is to understand the puzzle in explanatory terms and argue that principles of explanation such as *abstraction* and *stability* may shed light on the distinction between effects and by-products (Franklin-Hall forthcoming-a).[140] Such an account still targets an objective difference between effects and by-products, but not a metaphysically deep one. Following Franklin-Hall's discussion of causal explanation (but applied on the effect side), the mentioning of effects, like that of causes, is sparse and systematic. It is *sparse* in that we never mention most or the whole set of consequents deriving from a cause or cluster of causes. It is *systematic* in the sense that effects emerge with a certain regularity, which indicates concealed principles dictating their selection.

To characterise sparseness, we need a restrictive principle, working through censorship *via* procedures like deletion and abstraction. In describing a certain effect, we mark it off through the deletion of some of its particular causal consequences, or, in cases of abstraction, we may describe it in less detail through operations such as coarse-graining, amalgamation, or populational transformation. Franklin-Hall explains in detail:

> In coarse-graining, a particular feature is described as falling in some range or exceeding some threshold (*over 30 miles per hour* replacing *35 miles per hour*). In amalgamation, multiple lower-level features applicable to a particular individual are combined in a more complex parameter, and the particular values of the components are thereby lost (*15 kgm/s momentum* replacing *5 kg mass at 3 m/s velocity*). In populational transformation, a population-level feature is cited rather than a set of parameters applicable to the individuals constituting the population (*temperature* of a gas, e.g., mean kinetic energy, replacing its constituent molecule's particular *kinetic energies* (themselves products of amalgamation, since kinetic energy ≈ 1/2 mass * velocity)). (…) [A]ll of these kinds of transformation are applied in different orders and degrees to the complete story for a particular event. (Franklin-Hall forthcoming-a: 14, emphasis in original).

To characterise systematicity, we need a principle tracking stability, that is, the extent to which the target effect would have still occurred had the circumstances been different. The idea of stability is understood by appealing to a possible worlds framework, *viz*. the stability of a certain event is equivalent to the set of privileged close possible worlds (only slightly differing from our own) in which the target event nevertheless takes place. For instance, an event such as WWI is stable in the sense that

[140] Of course, the previous relativisation strategy also works in explanatory terms (with the choice of reference frames depending on the explanatory project pursued), however, since relativisation appears to be problematic, the focus will be on Franklin-Hall's more recent 'explanatory economy' account. Franklin-Hall (forthcoming-a: 9-11) acknowledges the limits of relativisation views (frameworking and populational) and takes her account to be theoretically preferable.

it would have taken place even if lots of other things had been different (for a more detailed account, see Franklin-Hall [forthcoming-b]).

The combination between principles of abstraction and stability generate a more general, *economical* formula: effect projections 'cost less' in virtue of being sparse (abstract or narrowed down) and 'deliver more' in virtue of showing stability or robustness. In her discussion of causal explanation, Franklin-Hall calls this combination the Causal Economy principle or the 'biggest bang-for-your-buck standard' (Franklin-Hall forthcoming-a: 12). Its main motivation is directly related to our limited cognitive capacities. Complete explanations are not within our grasp, so in order to get good explanations we need to appeal to a combination of abstraction and stability. By analogy, complete effect projections are unlikely, so in order to have robust effect projections we need a related principle for the effect side. It is a straightforward exercise to apply a similar economy principle on the effect side: it is 'cheaper' not to mention the entire complex architecture of by-products deriving from an occurrence and more 'profitable' to mention effects with a high recurrence, that is, effects that would have occurred even if lots of other things had been different.[141]

2.4 Is There an Explanation-Based Difference Between Effects and By-Products?

The initial temptation is to accept explanatory principles as marking the difference between effects and by-products, however, there are important reasons to be reserved about an explanation-based resolution to selection problems.

First, it is not clear if selection puzzles are best framed in explanatory terms. A blunt objection is that the selection of relata is a fundamental issue in the metaphysics of causation, on a par with essential and controversial features like connection or direction, rather than a problem specific to accounts of explanation. Since all particular causal claims already select, we would not have a grasp on the notion of causation without selection. If it appears as a problem of explanation it is because it is first a problem of causation (given that causal explanations are grounded in the causal relation[142]). In reply, several authors choose to downplay the difference between

[141] I am glossing over a large number of features of Franklin-Hall's view (but see Franklin-Hall forthcoming-b for a detailed account), as I am only interested if it may be extended to offer an explanation-based resolution of the effect selection puzzle.

[142] For a detailed defence of explanatory realism – the view that a cause C explains an effect E in virtue of an objective relation R relating events or states of affairs in the world and R is the causal relation – see Kim (1981, 1993d).

causation and causal explanation in order to be able to offer problems of causation a treatment informed by the discussions specific to theories of explanation (van Fraassen 1980, Woodward 1984, Bennett 1988, Hitchcock 1996). For instance, if causation traditionally describes a relation between events and causal explanation traditionally describes a relation between propositions, Woodward (1984) argued that causal explanation may relate events, while Bennett (1988) argued that causation may relate propositions. But in this context such a reply is misleading. With the selection of causes vs. conditions and of effects vs. by-products the problem is not one of liberalism about the nature of relata, but one of allowing the shifty character of explanation to characterise the distinctions we want to turn out objective. If the selection of causes vs. conditions and of effects vs. by-products is guided by explanatory aims and explanation is guided by particular interests, then selection will turn out interest-dependent (subjective and capricious rather than objective and stable).[143]

Second, it is not clear that explanatory principles could deliver the sought objective difference between effects and by-products. One may attack the concept of economy as being a mind-dependent construct, a formula encompassing several more or less reasonable notions (fairly abstract character, stability, robustness, sufficiency, etc). However, this would set the bar of objectivity too high and ultimately show an uncharitable approach to effect selection. In contrast, a more serious concern is that for a particular occurrence there could be more than two economical packages costing little and delivering a lot. Recall Collingwood's (1940: 304) example of variation in the description of causal facts. A car corners too fast, skids and crashes: the driver takes the cause of the accident to be her cornering too fast, the road engineer takes the cause of the accident to be the defective road surface, and the car manufacturer takes the cause of the accident to be the defective design of the engine. Complex examples showcase the presence of competing causal explanations and the fact that none wins the causal economy race. With complexity in, all explanations look partial and again, either conversational pragmatics or relativisation to reference frames would be needed to select amongst options. Does this undermine the economy account? Franklin-Hall thinks not, mainly because in most causal systems the winners of the explanatory race will be 'relatively singular rather than a combination of factors' (Franklin-Hall

[143] Of course, one may bite the bullet and take causation itself to be interest-dependent, however, that entails renouncing the idea that selection is not arbitrary.

forthcoming-a: 25). But this view is inaccurate. On a closer look, any causal system is a highly complex system in which a combination of causal factors compete. Relatively singular, more abstract factors will win the race only if one assumes that a causal system can be described in sufficiently abstract terms, however, abstraction should be a part of the economical solution rather than a general presumption about causal systems.[144] Again, it is a straightforward exercise to extend the analogy to effect selection. By the same token, we would be faced with a choice amongst equally relevant consequents and need to decide what counts as a genuine effect and a mere by-product in complex cases with seriatim and parallel effects. If in such cases an array of consequents compete for the status of effects, considerations of economy alone cannot provide a clear differentiation between effects and by-products.

2.5 Summing Up

Effects and by-products appear to be different in an important sense. I have started from the idea that a prospective account of effect selection should look for a distinction between effects and by-products that is neither metaphysically deep, nor interest-dependent. I considered that a natural methodological move is to import the strategies relevant on the cause side and apply them on the effect side: both selection problems belong to the same relation and, if transitivity holds, they should be in principle highly similar. I have examined two different ways of approaching selection, relativisation (Hesslow 1988, Cheng and Novick 1991, Schaffer 2005, Waters 2007) and explanatory economy (Franklin-Hall forthcoming-a). Both strategies fail when applied on the effect side, and as a result, it is unclear what actually guides the distinction we draw between effects and by-products. In fact, the distinction seems highly arbitrary. Lastly, it may be that effect selection cannot even receive a similar treatment to causal selection, as their relation is not symmetric. Here are several important dissimilarities:

(1) Effect selection means 'choosing' or 'perceiving' a certain target event against parallel or downstream consequents, whereas causal selection means contrasting causes with conditions. Some selection principles apply on both sides, but they seem to be different. As Schaffer (2005: 352) remarks, effect selection seems to be related to

[144] Indeed, even a supporter of the relativisation strategy needs to start from a certain level of abstraction. At this point, one could insist that abstraction concerns vertical selection, while the idea of relativisation represents an answer to the genuine horizontal problem of selection. Alternatively, one could admit that the best theory should ideally strike a balance between economy and relativisation.

agential intentions rather than associated with the contrastivity of causal inquiry. However, it is one thing to claim that interest guides the choice of target events and a different thing to claim that interest marks the distinction between effects and other parallel or downstream by-products. If transitivity holds and effects become causes, they should in principle receive a similar treatment. But how could they if interest reigns only over the effect side? [145]

(2) There seem to be more selection criteria governing causal selection than effect selection: for instance, sometimes causal selection can be guided by the predictive value of certain causes (certain causes get selected because they predict an effect most effectively). In contrast, effect selection cannot be guided by predictive value. Other criteria of causal selection – but not of effect selection – include variability, deviation from a theoretical ideal, replaceable character, and necessity (for a detailed list of causal selection criteria, see Hesslow 1988).

(3) Lastly, the distinction between causes and conditions is general and straightforward – it is just the distinction between 'active' difference makers and existing circumstances, to use different terms. The distinction between effects and by-products is not straightforward; on the contrary, it is difficult to create an accurate complex map of different types of by-products and extract the main concepts dictating their specificity.

A rapid look at the linguistic evidence shows a significant diversity of forms. As a general remark, the problem of effect selection starts from the simple thought that something is a by-product if it has a secondary nature, but then things get complicated.

An initial distinction is drawn between effects and by-products related to certain states of an *agent*, e.g., unexpected, unforeseen, unintended, or incidental effects. For instance, while self-knowledge may be philosophically attractive, it may have unforeseen effects such as being emotionally sobering or outright depressing. Alternatively, an academic career may be conducive to a life of reflection, but it may have unexpected effects such as dealing with competition or hard egos.

[145] And not only if transitivity holds. Causes and effects are indistinguishable in cases of two-way causation (e.g., Twitter activity and TV viewership: do tweets cause TV watching, or does TV watching cause Twitter chatting?).

A second distinction is drawn between effects and by-products related to the idea of *purpose*. In a sense, something is deemed a by-product if it emerges in the purpose-oriented process of producing (or taking apart) something else. For instance, molasses are a by-product of sugar refining, straw is a by-product of grain harvesting, and at a more abstract level, poverty can be seen as a by-product of colonial prosperity. In a related sense, something is called a side effect if it is not a purpose-specified result: if the purpose of loratadine is to alleviate allergy symptoms, effects such as drowsiness, dizziness, or blurred vision will be described as side effects. Side effects may be specified along further dimensions – for example, the side effects of a prescription drug are specified according to their frequency (common effects, uncommon effects, rare effects, very rare effects). Lastly, a particularly interesting notion is that of an end result, which seems to be built on a certain purpose or function. For instance, a space station is an end result of years of research; alternatively, the functions of the human brain are the end result of billions of years of evolution.

A third distinction is that between effects and by-products related to a *temporal dimension* (or spatio-temporal to be completely accurate). Linguistic evidence frequently indicates a distinction between effects and remote effects, after-effects, or long-term effects. In addition, time is involved in the idea that causation has a scalar nature and effects peter out and disappear like ripples on a pond after a stone has been thrown in it (cf. Chapter III, Section 3.2). Lastly, a temporal dimension is to some extent involved in the discussion of hypothetical effects in the modelling of different situations, e.g., climatic scenarios following asteroid impacts or countervalue nuclear war. For instance, we know that a remote climatic effect of a possible countervalue nuclear war is an atomic winter.

The presence of these asymmetries, the taxonomical complexity and the failure of relativisation and explanatory economy combine to generate uncertainty about effect selection. Despite the strong sense that effects and by-products are different, the criteria governing their differentiation are neither clear, nor stable.

3. Effect Selection and Responsibility for Consequences

Given that we are morally responsible not only for what we do, but also for the consequences following from our acts, which of those consequences are we responsible for and why? To consequentialist theories, the morality of an act is exhausted by the moral status of its consequences. Relevant consequences can be causal (what causally follows from our acts), logical (what is entailed by an act, e.g. killing entails death, but does not cause it), both (a view defended for instance by Sosa [1993]), or may include the whole state of the world following the act (a position taken by Feldman [1997]). Focusing on causal consequentialism (and bearing in mind the network model), one notices that acts have far too many consequences, more than could matter to their moral status. Out of the things we cause, how do we establish which ones matter morally and which not? Is there a principle limiting the countless number of consequences an agent could be responsible for? In this section, I use the concerns about appropriate effect selection to formulate an epistemic argument against consequentialism. The problem is briefly this: given the radical uncertainty about the criteria governing the distinction between effects and by-products, there seems to be no clear difference between the relevant and the irrelevant consequences of a given act, so when it comes to assessing the moral quality of an act as a function of its consequences, we are not in a position to know what they really entail. For instance, if an act has immediate good consequences and remote bad consequences, what exactly takes priority in the moral assessment of the act if there is no clear explanation of the priority typically given to a consequence or set of consequences (*the Effect*) over other results (*by-products*)?

3.1 Epistemic Arguments

It is useful to compare the problem I have raised (a selection-based objection) with the classic epistemic argument against consequentialism (Kagan 1998, Lenman 2000). For instance, Kagan notes:

> Perhaps the most common objection to consequentialism is this: it is impossible to know the future. This means that you will never be absolutely certain as to what all the consequences of your act will be. An act that looks like it will lead to the best results overall may turn out badly, since things often don't turn out the way you think they will: something extremely unlikely may happen, and an act that was overwhelmingly likely to lead to good results might-for reasons beyond your control-produce disaster. Or there may be long term bad effects from your act, side effects that were unforeseen and indeed unforeseeable. In fact lacking a crystal ball, how could you possibly tell what all the effects of your act will be? So how can we tell which act will lead to the best results overall-counting all the results? This seems to mean that consequentialism will be unusable as a moral guide to action. All the evidence available at the time of acting may have pointed to the conclusion that a given act was the right act to

perform-and yet it may still turn out that what you did had horrible results, and so in fact was morally wrong. Indeed, it will never be possible to say for sure that any given act was right or wrong, since any event can continue to have further unseen effects down through history. Yet if it is impossible to tell whether any act is morally right or wrong, how can consequentialism possibly be a correct moral theory? (Kagan 1998: 64).

The classic epistemic argument against consequentialism receives a compelling formulation in Lenman's (2000) as the 'argument from cluelessness': since most downstream consequences following from our acts are inscrutable, when it comes to assessing our acts as a function of their consequences we are clueless. Since consequentialism is the view that the rightness of an act is a matter of its good overall consequences and we seem to be clueless about a significant percentage of them, it follows that we cannot assess acts as a function of good overall consequences and therefore consequentialism fails. As Lenman remarks,

The worry is (…) that we do not have a clue about the overall consequences of many of our actions. Or rather – for let us be precise – a clue is precisely what we do have, but it is a clue of bewildering insignificance bordering on uselessness (…). We may not be *strictly* without a clue, but we are *virtually* without a clue. The trouble for consequentialism then is that the foreseeable consequences of an action are so often a drop in the ocean of its actual consequences. (Lenman 2000: 349-350, emphasis in original).

In comparison, my epistemic argument from effect selection is similar in aims but different in accent. With effect selection, the uncertainty is not as much about being clueless regarding the distant future consequences of our acts, but about not having a clear idea of what relevant consequence(s) take priority in the moral assessment of a particular act. If the cluelessness argument insists that all results determine the moral status of an act (immediate effects, side effects, remote consequences, indirect consequences etc.), the effect selection argument insists that we do not know why some results rather than others determine the moral status of an act (all the more if some consequences are good and other consequences bad). If my analysis of effect selection is correct, the distinction between significant and irrelevant consequences of our acts is arbitrary and evasive, so how could they legitimately and accurately define the rightness of an act? In more formal terms:

(1) Because consequentialism determines the moral status of an action φ by its consequences, one has access to the moral status of an action φ if one has access to its consequences.

(2) In terms of access to consequences, consequentialists will need to distinguish between the 'primary' effects which are determinative of an act's rightness and those 'secondary' effects which are not. (One reason is that anyone thinking about effects must

engage in effect selection; another reason – which I will come to in the next paragraph – is that drawing this distinction might seem to offer a way for consequentialists to overcome the argument from cluelessness).

(3) But because one does not know what the relevant or primary consequences of a particular action φ are (in contrast to its irrelevant or secondary consequences), one does not know that action's moral status.

(4) But if one cannot know the moral status of our actions (the comparative quality of their consequences), then one does not know which actions consequentialism requires one to perform.

(5) But if one wants consequentialism to offer significant moral guidance to agents, agents must know which actions consequentialism asks them to perform.

(6) Therefore, consequentialism cannot offer us significant moral guidance.

Alternatively, the argument from effect selection could be framed not in contrast to the cluelessness argument, but as a reaction to it, *viz*. if consequentialists are to avoid the problems arising from the fact that each action has countless consequences, they will need a way to distinguish relevant from irrelevant consequences or primary from secondary consequences so as to evaluate the morality of an act. But, the argument goes, there is no principled way to distinguish relevant from irrelevant consequences or primary from secondary consequences because the criteria governing effect selection are neither clear, nor stable (*as per* Section 2). If there is no clear explanation of the priority given to a set of consequences over other results, then there is no principled way to evaluate the morality of acts as a function of their consequences and therefore consequentialism fails. In different terms, there is simply no non-arbitrary way of partitioning the space of consequences such that some of them to determine the status of the act.

An immediate worry is that a consequentialist's distinction between relevant and irrelevant consequences is unlikely to map onto our intuitive distinction between effects and by-products. This is because most consequentialists would allow for certain things we intuitively call by-products or side-effects to nonetheless be relevant to the moral assessment of an action. Thus whether to engage in a particular act will depend not only on the 'goodness' of its consequences but also on what we intuitively call its side-effects. For example, the currently planned American airstrikes on Syria may be taken to have good effects (confronting extremism in the Middle East) and bad side-effects

(civilian deaths, propping up the regime of Syrian dictator Bashar al-Assad, tension in the region etc.), but on balance a consequentialist assessment of them is deemed positive. Both consequences and side-effects, good and bad, are weighed up in reaching a consequentialist assessment of the action. But even if the consequentialist somehow avoids having a vast number of consequences relevant to the moral quality of a given action, she cannot avoid all selection criteria, that is, she will still need to distinguish relevant from irrelevant consequences. If no such distinction is available, it is difficult to assess the moral status of an act. Two further clarifying remarks are in order.

First, both epistemic arguments against consequentialism work only against what Lenman (2000: 343) calls 'pure consequentialism'. Pure consequentialism is the view that a person should be concerned with the rightness of her acts as a matter of the goodness of all its consequences – short term effects, long term effects, side effects, indirect effects, all of them determine the moral status of a certain act. At the opposite end, impure consequentialism is the view that a person should act such that if (almost) every other person acted in the same way the good would be maximally promoted. Against the doctrine of pure consequentialism, the classical argument from cluelessness introduces a *quantitative* concern about the potentially infinite number of relevant effects. In contrast, the effect selection argument introduces a *qualitative* concern about the status of effects (relevant vs. irrelevant, primary vs. secondary). However, it must be noted that even if consequentialists managed to isolate an area of relevant proximate effects (so as to deal with the quantitative concern), any simple act would still entail a large number of effects and by-products, a multiplicity extending not forward in time, but immediately from a given action. Even in such cases the consequentialist will need some selection criteria. Therefore, if it is argued that for moral reasons we should deem acts right or wrong on the basis of a restricted area of consequences (e.g., the results that we are aware of, intend, or able to foresee), one may reply a) that epistemic arguments work against a pure version of consequentialism, and b) that even a restricted area of consequences will still include a significant number of effects and by-products, and c) that foreseeability includes an awareness of potential causal ramification.[146]

[146] More clearly, it is not possible to say that we select the effects we are able to foresee if most of us are aware or able to foresee that causal chains will ramify endlessly. In different terms, one cannot restrict effect selection to certain foreseeable consequences if one is aware of the endless number of foreseeable causal consequences.

Second, both epistemic arguments against consequentialism would be better understood if we clarify a missing premise regarding transitivity. For the secondary or remote consequences to count in the assessment of our acts, transitivity must hold. Naturally, a causal consequentialist would attempt to block the conclusion that we are responsible for every consequence following from our acts. As Sartorio (2009b) points out, one way is to invoke the legal notion of proximate causation and argue that an agent is a cause of an outcome only if the outcome directly derives from her agency. On this view, other remote or secondary effects fail to count as genuine consequences. In addition, if other agents or unexpected events intervene, causal chains break and the initial agent is no longer causally responsible. However, what counts as 'proximate' in proximate causation is something heavily dependent on norms and expectations, a fact that makes proximate causation incompatible with the general idea that causation is an objective and extensional relation (for a detailed critical discussion of this core legal notion, see for instance Moore 2009). Another way to block the conclusion that we are responsible for every consequence deriving from our acts is to reject transitivity altogether. If transitivity does not hold, then some connections in the endless causal chains may not be causal; therefore, not every consequence following from one's act will influence the moral status of that act.[147] However, transitivity could be rejected only in some cases. To reject transitivity means showing that transitivity fails *sometimes*. It does not mean that if A causes B and B causes C, A *never* causes C.[148] And even if transitivity fails only sometimes, it is not very clear what separates the cases in which it fails from the cases in which it does not. For instance, if an agent shoots at a person and that person ducks and survives, the shooting causes her to duck and the ducking causes her to survive, however, we do not say that the shooting causes the survival. In addition, it is not perfectly clear what explains the failures of transitivity (but see Broadbent [2010] for an explanation based on reverse counterfactuals of the form 'If c causes e then $\neg Oe \Box\!\!\rightarrow \neg Oc$').[149] Ascending a level, it may be that all failures of transitivity are nothing but mere psychological effects. For instance, if some find implausible the

[147] Here is an example due to Ned Hall (2004): Agent A spots a boulder bouncing towards him, ducks and survives. The boulder coming towards him causes him to duck, and the ducking causes his survival. But we do not say that the boulder bouncing caused his survival.

[148] Non-transitivity is different from intransitivity. A relation R is non-transitive iff $\neg\forall(x, y, z)((Rxy \& Rxz) \supset Rxz)$ and intransitive iff $\forall(x, y, z)((Rxy \& Rxz) \supset \neg Rxz)$ (see Broadbent [2010] for a detailed analysis).

[149] Reverse counterfactuals – had the effect not been present, the cause would not have been present – can show how transitivity fails, e.g., it does not make sense to say 'had the survival did not occur, the ducking (or the shooting) would not have occurred'.

thought that a lack of a nail may cause the loss of a kingdom,[150] one may wonder whether the links in the sequence are really causal or we deal with a pattern of difference-making which depends on external considerations.

As a last remark, both versions of the epistemic argument may be combined to generate a more compelling contention against consequentialism: not only that we are clueless about the overall consequences of our acts, but we are clueless about which consequences are actually relevant for the moral assessment of our acts. If an epistemic objection based on effect selection fails to be convincing on its own, it may at least make the classic epistemic argument more powerful. If cluelessness goes even deeper, we end up with very weak reasons to think that our actions are right or wrong by consequentialist principles.

3.2 Several Objections

To showcase the seriousness of the problem I have raised, I will next test the selection-based version of the epistemic argument against a battery of objections usually faced by the classic 'cluelessness' version.

(1) Scalarity

The scalarity response to the classic epistemic argument is that one should not grant much significance to distant consequences, as they tend to peter out after a time. Both causal relations and moral blameworthiness seem scalar notions, that is, matters of more-or-less which allow a comparison of degrees rather than matters of either-or. Several theorists explicitly endorse scalarity. For example:

● 'The effects of any individual action seem, after a sufficient space of time, to be found only in trifling modifications spread over a very wide area, whereas its immediate effects consist in some prominent modification of a comparatively narrow area'. (Moore 1903: 153).

[150] For instance: 'For want of a nail the shoe was lost; for want of a shoe the horse was lost; for want of a horse the rider was lost; for want of a rider the battle was lost; for want of a battle the kingdom was lost – and all for the want of a horseshoe nail'. (Broadbent 2010: 18).

● '[Remote consequences] approximate rapidly to zero like the furthermost ripples on a pond after a stone has been dropped on it'. (Smart and Williams 1973: 33).

● '[Causation is a] matter of continuous variation. We live in an analogue world, a world more accurately depicted by the delicate shadings of water colours than by the bright lines of mosaic tiles (Quine 1960: 127)'. (Moore 2012: 446).

Admittedly, the scalar nature of causation is a controversial issue (cf. Chapter 3, Section 3.2). However, things become clearer if attention is focused on the causal ramification in dynamic systems with high sensitivity to small variations in initial conditions (Lenman 2000). For instance, we cannot possibly know the extremely widespread differences made by financial markets to people's plans, states of mind, or safety, but at the same time it would be absurd to claim that the effects flowing from a certain dynamics of financial markets fail to reach individuals' lives because they peter out in time. It is precisely in this kind of systems that one cannot avoid taking into consideration remote consequences (as any small-scale action, e.g., a particular financial transaction, will ramify massively). But a defender of the classic epistemic argument could accept the challenge and focus on the immediate effects of an act, *viz*. those consisting in some 'prominent modifications'. It is in this case that a selection-based version of the epistemic argument becomes relevant. How exactly is a 'prominent modification' delineated and what makes others less prominent? And if several such immediate consequences are generated, which one should determine the moral status of an act? What if they turn out to be qualitatively different? Without getting into too many details, the general idea is that problems of effect selection may consolidate the classic epistemic argument.

(2) Cancelling-Out

Another potential move to be made against the classic epistemic argument is Kagan's 'cancelling out' response:

> [I]t remains true that there will be a very small chance of some totally unforeseen disaster resulting from your act. But it seems equally true that there will be a correspondingly very small chance of your act resulting in something fantastically wonderful, although totally unforeseen. If there is indeed no reason to expect either, then the two possibilities will cancel each other out as we try to decide how to act. (Kagan 1998: 65).

The thought here is that totally unforeseen and remote good and bad consequences will somehow cancel each other out allowing the foreseen consequences to count. Supporters of the cluelessness argument could argue against the existence of obvious empirical or a priori probabilistic grounds to assume the cancelling-out hypothesis (see Lenman [2000] for a detailed discussion), that is, even if such a hypothesis looks plausible there is no evidence for it. However, a theorist who relies on the effect selection argument could accept the cancelling-out idea (at least as far as remote consequences are concerned). For I have mentioned that even if consequentialists manage to deal with the quantitative concern and focus on an area of relevant proximate effects, any act would still generate a multiplicity of effects and side-effects requiring some selection criteria. In addition, one may ask if the cancelling-out hypothesis does not apply in this restricted area of consequences – after all, when is causal ramification complex enough so that good and bad consequences will cancel each other out?

(3) Actual vs. Expected Consequences, Rightness vs. Decision-Procedure

Another popular objection to classic epistemic arguments (Jackson 1991, Railton 1984, Kagan 1998) goes as follows: it does not matter that a significant number of consequences are unforeseen because the moral status of an act is not determined by the whole set of actual, objective consequences flowing from that act, but only by the expected, subjective consequences a reasonable agent could envisage. But one problem is that we need to answer why we should take into account subjective expected consequences and the simple answer is that in many morally relevant cases the expected consequences are simply means of approximating objective rightness. If we do not know how objectively right an action will be – either because we are clueless about its remote consequences, or because we cannot distinguish between its relevant and irrelevant consequences – then the objection fails. If the expected consequences are not means of approaching objective rightness, then we could not possibly assess the moral status of an act.

According to a different, more refined version of this objection, consequentialism should provide a general criterion of rightness rather than a decision procedure – that is, it should be a principle about the moral status of our acts rather than

about the way we come to know and understand what the morally right act is.[151] But the two must be somehow related, in the sense that a theory endorsing a criterion of rightness should also generate an action-guiding rule. If it generates such an action-guiding rule, then epistemic arguments become relevant again (as we should govern our conduct in the light of what we know about the consequences of our acts). If it does not, then we end up with a sort of *disengaged consequentialism*, perhaps true, but useless in this world (Lenman 1998: 360-361).

(4) Companions in Guilt I (Kagan)

Another objection against the cluelessness argument takes the companions in guilt form:

> [The Epistemic Argument] threatens not only consequentialism, but indeed all plausible normative theories. For if it is in fact impossible to get a grip on the consequences of an act, then this problem will be inherited by all theories that give this factor any weight at all and this will be virtually all theories, For... all plausible theories agree that goodness of consequences is at least one factor relevant to the moral status of acts. (Kagan 1998: 64).

In reply, it is worthwhile to remind the reader that epistemic arguments make sense against pure consequentialism and not all moral theories out there give the same weight to *all* consequences following from an act. Non-consequentialists may simply choose to ignore remote consequences as morally irrelevant. On the other hand, the way we individuate consequences is indeed a concern for virtually all normative theories granting them any weight. For not having clear and stable principles of effect selection makes it difficult to get a grip on the consequences that matter. Therefore, while the cluelessness argument cannot be meaningfully extended to affect non-consequentialist theories, the effect selection argument could in principle affect all plausible normative theories taking consequences into account.

(5) Companions in Guilt II (Dorsey)

A similar objection against the cluelessness argument goes as follows:

> I shan't dispute the claim that many of the consequences of our actions are unknowable to us. The question I shall address is whether the fact of cluelessness should worry the advocate of consequentialism. I claim it should not. In particular, I argue that there is very good reason to believe that the problem of cluelessness for consequentialism is only as embarrassing for consequentialism as the spectre of epistemological skepticism is embarrassing for metaphysical realism. But coping with epistemological skepticism rarely

[151] Lenman (2000: 360) remarks that such a move is revisionist in that it maintains consequentialism as an account of objective rightness while at the same time advancing a non-consequentialist story about subjective rightness.

tempts anyone to abandon metaphysical realism. Moral philosophers should face no greater temptation when it comes to consequentialism. (Dorsey 2012: 49).

One may be immediately tempted to answer that this is not much of an argument. After all, a particular position is not made true by the fact that the philosophers supporting it fail to give in to certain temptations. It may be that metaphysical anti-realism is true, and if it is, its truth would be completely independent from whatever temptations metaphysical realists happen to disregard.[152] The same goes for consequentialism and its foes. On a different note, if one takes scepticism seriously, the options are (a) to accept the truth of scepticism but continue to support metaphysical realism anyway, (b) to accept the truth of scepticism and deny metaphysical realism, or (c) to hold that sceptical challenges affect consequentialism but not metaphysical realism (Dorsey 2012: 51). I am happy to allow option (c) so that the sceptical argument from analogy goes through. According to option (a) we accept that scepticism is true but continue to work with the things in the world as they appear to us (Norcross 1990). In Dorsey's terms, we 'soldier on' and 'make do with what we have' (Dorsey 2012: 58). Two remarks are in order: first, 'soldiering on' may be interpreted in a positive key as a way of dealing with the sceptical challenge, in the sense that when we make do with what we have we leave sceptical worries aside. By extension, consequentialism could be defended as a view that makes do with the consequences we know about while leaving aside epistemic worries about remote consequences. However, if 'soldiering on' bypasses the argument from cluelessness, it cannot avert the epistemic argument from effect selection, which affects precisely the comparable quality of the consequences we know about. Second, 'soldiering on' may also be interpreted in a more negative key as a way of acknowledging the stringency of the sceptical challenge. This interpretation leads to the last option, (b), accepting the truth of scepticism and rejecting metaphysical realism. By extension, this option implies acknowledging the seriousness of the epistemic arguments and holding that consequentialism fails as a result. While I shall not argue for or against metaphysical realism as an answer to scepticism, it is worth pointing out that one may at least cast doubt on metaphysical realism about *causation* (cf. Chapter I), which seems to be especially relevant in this context both to consequentialism and to the epistemic arguments against it. More clearly, if there are

[152] And why would they not disregard or look to downplay the strength of arguments supporting the opposite view anyway?

some reasons to take seriously certain concerns about causal realism, perhaps one should be similarly worried about causal consequentialism.

If my epistemic argument against consequentialism survives these objections, we end up with weak reasons to think that our actions are right or wrong by consequentialist principles.[153]

4. Conclusions

In this chapter, I examined the problem of effect selection and argued that a prospective account will need to strike a balance between two unappealing extremes: (a) the claim that effects and by-products are metaphysically distinct, and (b) the claim that there is no sense in which effects and by-products are objectively different and selection is always governed by context-dependent pragmatics. I argued that despite the strong sense that effects and by-products are essentially different, the criteria governing their differentiation are not clear or predictable. Furthermore, I used the results gleaned from the first sections of the chapter to develop an epistemic argument against consequentialism. I defended this argument against several objections.

[153] The success of consequentialism may not be related to the number or comparable quality of consequences, but to the clarification of our status as moral agents within complex causal chains. On the one hand, we have no problem to describe physical effects as reverberating as far as the laws of the universe allow: 'an explosion may cause a flash of light which will be propagated as far as the outer nebulae' (Hart and Honoré 1985: 68). On the other hand, we seem reluctant to treat actions similarly (at least sometimes). As Nagel famously remarked, the idea of agency is incompatible with actions and people being part of natural causal networks, '[b]ut as the external determinants of what someone has done are gradually exposed, in their effect on consequences, character, and choice itself, it becomes gradually clear that actions are events and people things. Eventually nothing remains which can be ascribed to the responsible self, and we are left with nothing but a portion of the larger sequence of events, which can be deplored or celebrated, but not blamed or praised'. Nagel (1979: 37).

Summary and Conclusions

My dissertation examined different ways in which causation and responsibility intersect. I argued that some central and controversial features of causation are relevant to the understanding of certain moral notions and views. Structurally, the dissertation was organised into four self-contained but loosely interconnected chapters centred on how the causal and the moral overlap. The project was motivated by the presence in the literature of several aspects of causation that are underresearched in relation to ethical theory and are consequential in two respects: a) they dispute assumptions or features of widely accepted views of responsibility; b) they offer additional support to certain moral perspectives. The specific objective of the project was to investigate the extent to which an appeal to causation explains moral responsibility attributions and increases the strength of some specific moral views.

I started from the straightforward intuition that responsibility rests in important respects on causation. Since the only way an agent can make a difference in the world is through her causal powers, causation grounds responsibility in the sense that it determines, explains or makes true responsibility claims. This basic grounding relation is a recurrent theme throughout the dissertation.

In Chapter I, I examined the realist assumption that causation is an objective, extensional relation between space-time located relata. A realist stance about causation is meant to offer moral evaluation a naturalistic basis and make moral properties continuous with a naturalistic view of the world. The main contribution of the chapter was to show that such realist sympathies are problematic and by extension so are the views hoping to tie responsibility assessment to an objective, determinate relation in the world.[154] I am persuaded by the thesis that a single concept of causation is at work in metaphysics, morality, and law (Moore 2009, Schaffer 2010) and confident that metaphysical debates on causation will shed light on and decide disputes about core moral and legal concepts. However, I argued that causal relations may fail to provide moral assessment with the naturalistic basis hoped for, which may ultimately prove to be an important limitation of an otherwise promising research programme. Indirectly, my conclusions contribute to the debates about the objective reality of causation, siding with the critics of causal realism (Hume 1739, Russell 1913, Wittgenstein 1922, Putnam

[154] See, for instance, Moore (2009: 4-5).

1984, van Fraassen 1980, Blackburn 1993, Spohn 1993, Esfeld 2009). They also offer indirect support to the legal theorists who argue that causation in legal contexts differs from causation outside the law (Edgarton 1924, Malone 1956, Kelman 1987, Stapleton 2009) – although I am not convinced by this contention.

As the criticism of causal realism is an extensive topic, my strategy was to show that causation is based on incompatible intuitions and the best attempts to explain them fail to safeguard a robust sense of realism. I examined two such attempts: one defends a sense of objectivity through a reconciliation of our incompatible intuitions; the other aims to retain a different sense of objectivity within the structural equations framework (Spirtes, Glymour, and Scheines 1993, Pearl 2000, Halpern and Hitchcock 2010, 2013). I argued that both describing and modelling of causal facts are affected by deep-rooted ambiguities, and as a result, causal realists need to retreat towards more modest commitments. In terms of contributions, the analysis of both attempts provided new insights. For instance, I indicated the ambiguous role played by the selection constraints involved in the description of causal facts (related to Strategy I) and showed that considerably more work is required to obtain systems of equations accurate and complete in their predictions and interventions (related to Strategy II). Although critical, my remarks about the justification of modelling choices are not meant to discredit the general value of the structural equations framework as a formal device allowing access to causal structure, but to invite further reflection on the adequate construction of apt causal models.

In Chapter II, I examined how the context sensitivity of causal claims impacts moral assessment in complex situations and argued that taking context sensitivity seriously generates important worries about ultimate moral responsibility. The context sensitivity of causal claims has received sustained attention in the literature (Mackie 1980, Lewis 1973, Bennett 1995, Hitchcock 1996, Woodward 2003, Maslen 2004, Menzies 2004, Swanson 2010, Schaffer 2012), but it has not been discussed in connection to the idea that moral responsibility claims rest on causal claims (Sartorio 2007, Driver 2008, Moore 2009). Bringing together the two bodies of work generates the main contribution of the chapter, which is a new, causation-based sceptical argument regarding the possibility of ultimate moral responsibility. Similar to other sceptical views (Strawson 1994, Rosen 2004, Sinnott-Armstrong 2006a), it is independent of whether determinism is true or not. Unlike these sceptical views, it does not claim that we cannot be morally responsible, in a fundamental sense, because of the

way we are (our cognitive and social make-up), but because of the complexity of the world and of the way our causal knowledge is structured. Looking at the context sensitivity of causal claims one becomes aware of the complexity of causal determinants anteceding any outcome of interest, and the more one knows about the causal complexity anteceding a particular event, the more uncertain one becomes about what actually causes it. The more one becomes aware of the minute specificity of the causal world, the more one begins to realise how difficult it is to explain things fully.

In Chapter III, I examined the concept of higher-level causation as it appears in relation to non-reductive views of group agency. In the last decade, a number of philosophers have sought to associate agency with groups and describe them as real and robust entities, irreducible to the sets of networked individuals they are constituted of (Pettit 2003, 2007, Tollefsen 2002a, 2002b, 2003, Tuomela 2005, 2007, Copp 2006, 2007, List and Pettit 2006, 2011, Gilbert 2009, 2013). Most such non-reductive views focus on the criteria required to elevate groups to the status of novel centres of judgment, intention and action, but show comparatively little interest in their actual causal efficacy as relatively independent entities. I argued that non-reductive accounts lack an adequate causal story about how group agents impact the world. My argument contributes to increasing the plausibility of individualist views (in the debate between individualists and collectivists) and extends to all potentially autonomous collective entities discussed in the social sciences.[155]

Amongst the criteria required for group agency, non-reductive theorists such as List and Pettit (2011) include representational states, motivational states, and a capacity to process these states and actively intervene in the world on that basis. My focus was on the idea that genuine group agents are required to play a causal role that is in a sense irreducible to the sum of individual causal contributions $c_1+c_2+\ldots c_n$ of all members. In different terms, non-reductive views need to test for a mode of action consistent with their non-reductive way of conceiving of agents. Obviously, any view discussing the

[155] Pettit offers several examples of groups entitled to autonomy: 'Instances (…) will be appointment and promotions committees; committees charged with deciding who is to win a certain prize or contract; trusts that have to make judgments on the basis of a trustee's instructions; associations or the executives of associations that have to justify their actions by reference to the group's charter; corporations that have to comply with policies endorsed by their shareholders; public bodies, be they bureaucratic committees or appointed boards, that have to discharge specific briefs; and governments that are more or less bound to party programs and principles. (…)Think of the political movement that has to work out a policy program; or the association that has to decide on the terms of its constitution; or the church that has to give an account of itself in the public forum; or the learned academy that seeks a voice in the larger world of politics and journalism'. (Pettit 2003: 172).

concept of higher-level causation will come in contact with concerns related to causal overdetermination. However, I argued that non-reductive theorists need a coherent causal story independently of whether the problems of overdetermination are decisive or not. I considered a reply based on the realisation-insensitivity of higher-level causal claims – since high-level properties have typically multiple low-level realisations, they cannot be identical with the low-level properties realising them and therefore possess causal powers independent from the causal powers of their low-level realisers. This line of reasoning is contentious because it ends up mixing dependence and production accounts of causation (Hall 2004), it tends to conflate explanatory role with real causal efficacy, and it leaves explanation and moral evaluation to bounce fortuitously between ontological levels. Lastly, I discussed the practical value of group agency views – particularly their pivotal role to social explanation, prediction and design. I pointed out some controversial assumptions behind non-reductive views, and offered a fictionalist recommendation concerning our stance towards group agency talk.

In Chapter IV, I offered an analysis of selection on the effect side of causal relations. If causes have infinitely many effects, but only one or a few are selectively highlighted and mentioned in causal claims, what determines their selection from the complete set of consequents? The main goal of the chapter was to explore the problem of effect selection and assess its relevance to determining the appropriate extent of responsibility for consequences. First, I argued that a prospective approach to effect selection needs to strike a balance between two unappealing extremes: (a) the claim that effects and by-products are metaphysically distinct, and (b) the claim that there is no sense in which effects and by-products are objectively different and selection is always governed by context-dependent pragmatics. I showed that despite the strong sense that effects and by-products are essentially different, the criteria governing their differentiation are not clear or predictable. Second, I argued that the difficulties in the way of a clear answer to the puzzle of effect selection motivate an epistemic argument against consequentialism. Given the radical uncertainty about the criteria governing the distinction between effects and by-products, there seems to be no clear difference between relevant and irrelevant consequences of a given act, so when it comes to assessing the moral quality of an act as a function of its consequences, we are not in a position to know what they really entail. In different terms, we do not know what exactly takes priority in the moral assessment of an act if there is no clear explanation

of the priority typically given to a consequence or set of consequences (*the Effect*) over other results (*by-products*) flowing from that act.

The chapter makes two contributions. First, it offers an analysis of effect selection.[156] Despite its negative findings, my discussion provides a needed beginning for a systematic investigation of the problems engendered by effect selection (for instance, an explanation of the asymmetries between causal selection and effect selection) with the goal of better understanding how they might affect extant theories of relata individuation, negative causation, or transitivity. Second, my new epistemic argument may add to the problems of consequentialism, or be combined with classical epistemic arguments (such as Lenman's [2000] cluelessness argument) to generate a more compelling contention against consequentialism.

To reiterate, my dissertation examined different contentious features of causation and their relevance to the understanding of certain moral notions and views. In particular, I argued that important questions arise about the objective reality of causal relations, the context sensitivity of causal claims, the idea of higher-level causation, and the mechanism of selection on the effect side of causal relations. Amongst other things, I derived sceptical conclusions about causal realism, ultimate moral responsibility, non-reductive views of group agency, and the principles of effect selection. In future research, I plan to use the conclusions I draw from the analyses in chapters II-IV to approach current problems with the notion of resultant moral luck, the individuation of effects of actions, and philosophical theories of explanation. My analysis of the context sensitivity of causal claims, the breakdown of higher-level causal efficacy, and the problems of effect selection provides a necessary starting point for a more refined understanding of our interaction with the world.

[156] There is currently no developed discussion of effect selection in the causation literature.

References

Achinstein, P. (1983). *The Nature of Explanation*. Oxford: Oxford University Press.

Alexander, S. (1920). *Space, Time and Deity.* London: Macmillan.

Alicke, M., Rose, D., and Bloom, D. (2011). Causation, Norm Violation, and Culpable Control. *Journal of Philosophy 108* (12), 670-696.

Alvarez, M. (2005). Agents, Actions and Reasons. *Philosophical Books 46* (1), 45-58.

Anderson, J. (1938). The Problem of Causality. *Australasian Journal of Philosophy 16* (2), 127-142.

Anscombe, G. E. M. (1993). Causality and Determination. In E. Sosa and M. Tooley (Eds.), *Causation* (pp. 88-104). Oxford: Oxford University Press.

Armstrong, D. M. (1999). The Open Door. In H. Sankey (Ed.), *Causation and Laws of Nature* (pp. 175-185). Dordrecht: Kluwer Academic Publishers.

Aronson, J. L. (1971). On the Grammar of 'Cause'. *Synthese 22* (3-4), 414-430.

Barker, E. (1915). *Political Thought in England from Herbert Spencer to the Present Day*. London: Williams and Norgate.

Barnes, E. (1994). Why P Rather Than Q? The Curiosities of Fact and Foil. *Philosophical Studies 73* (1), 35-53.

Baumann, P. (2011). A Puzzle about Responsibility. *Erkenntnis 74* (2), 207-224.

Beebee, H. (2004a). Chance-Changing Causal Processes. In P. Dowe and P. Noordhof (Eds.), *Cause and Chance: Causation in an Indeterministic World* (pp. 39-57). London: Routledge.

Beebee, H. (2004b). Causing and Nothingness. In J. Collins, N. Hall and L. A. Paul (Eds.), *Causation and Counterfactuals* (pp. 291-308). Cambridge: MIT Press.

Beebee, H. (2006). Does Anything Hold the Universe Together? *Synthese 149* (3), 509-533.

Bennett, J. (1988). *Events and their Names*. Indianapolis: Hackett Publishers.

Bennett, J. (1995). *The Act Itself.* Oxford: Clarendon Press.

Bernstein, S. (forthcoming-a). Overdetermination Underdetermined.

Bernstein, S. (forthcoming-b). Omission Impossible.

Björnsson, G. (2010) Consequentialism, Strawsonianism and the Illusion of Undermined Responsibility. Unpublished manuscript.

Björnsson, G. and Finlay, S. (2010). Metaethical Contextualism Defended. *Ethics, 121* (1), 7-36.

Blackburn, S. (1993). *Essays in Quasi-Realism.* New York: Oxford University Press.

Blackburn, S. (2005). Quasi-Realism No Fictionalism. In M. E. Kalderon (Ed.), *Fictionalism in Metaphysics* (pp. 322-338). Oxford: Clarendon Press.

Blanchard, T. and Schaffer, J. (forthcoming). Cause without Default. In H. Beebee, C. Hitchcock, and H. Price (Eds.), *Making a Difference*. Oxford: Oxford University Press.

Borgo, J. (1979). Causal Paradigms in Tort Law. *Journal of Legal Studies 8,* 419-455.

Botterill, G. (2010). Two Kinds of Causal Explanation. *Theoria 76* (4), 287-313.

Braham, M. and van Hees, M. (2009). Degrees of Causation. *Erkenntnis 71* (3), 323-344.

Braithwaite, J. and Drahos, P. (2000). *Global Business Regulation*. Cambridge: Cambridge University Press.

Bratman, M. (1993). Shared Intention. *Ethics 104* (1), 97-113.

Bratman, M. (2009). Shared Agency. In C. Mantzavinos (Ed.), *Philosophy of the Social Sciences: Philosophical Theory and Scientific Practice* (pp. 41-59). Cambridge: Cambridge University Press.

Broadbent, A. (2008). The Difference between Cause and Condition. *Proceedings of the Aristotelian Society 108* (1), 355-364.

Broadbent, A. (2010). Causes of Causes. *Philosophical Studies 158* (3), 457-476.

Brock, S. and Mares, E. (2007). *Realism and Anti-Realism*. Stocksfield: Acumen.

Brogaard, B. (2003). Epistemological Contextualism and the Problem of Moral Luck. *Pacific Philosophical Quarterly 84* (4), 351-370.

Brogaard, B. (2008). Moral Contextualism and Moral Relativism. *Philosophical Quarterly 58* (232), 385-409.

Bunzl, M. (1979). Causal Overdetermination. *Journal of Philosophy 76* (3), 134-150.

Cane, P. (2002). *Responsibility in Law and Morality*. Oxford: Hart Publishing.

Carroll, J. (1994). *Laws of Nature*. Cambridge: Cambridge University Press.

Cartwright, N. (1979). Causal Laws and Effective Strategies. *Noûs 13* (4), 419-437.

Cartwright, N. (2004). Causation: One Word, Many Things. *Philosophy of Science 71* (5), 805-819.

Castañeda, H-N. (1984). Causes, Causity, and Energy. In P. French, T. Uehling, Jr., and H. Wettstein (Eds.), *Midwest Studies in Philosophy 9* (1) (pp. 17-27). Minneapolis: University of Minnesota Press.

Chakravartty, A. (2005). Causal Realism: Events and Processes. *Erkenntnis 63* (1), 7-31.

Cheng, P. W. and Novick, L. R. (1991). Causes versus Enabling Conditions. *Cognition, 40,* 83-120.

Clark, A. (1994). Beliefs and Desires Incorporated. *The Journal of Philosophy 91* (8), 404-425.

Cohen, S. (1998). Contextualist Solutions to Epistemological Problems: Scepticism, Gettier, and the Lottery. *Australasian Journal of Philosophy 76* (2), 289-306.

Coleman, J. (1974). *Power and the Structure of Society*. New York: Norton.

Collingwood, R. G. (1937). On the So-Called Idea of Causation. *Proceedings of the Aristotelian Society 38*, 85-112.

Collingwood, R. G. (1940). *An Essay on Metaphysics*. Oxford: Clarendon Press.

Copp, D. (1979). Collective Actions and Secondary Actions. *American Philosophical Quarterly 16* (3), 177-186.

Copp, D. (2006). On the Agency of Certain Collective Entities: An Argument From 'Normative Autonomy'. *Midwest Studies in Philosophy 30* (1), 194-221.

Copp, D. (2007). The Collective Moral Autonomy Thesis. *Journal of Social Philosophy 38* (3), 369-388.

Corlett, A. J. (2001). Collective Moral Responsibility. *Journal of Social Philosophy 32* (4), 573-584.

Corry, R. (2006). Causal Realism and the Laws of Nature. *Philosophy of Science, 73* (3), 261-276.

Crane, A., Matten, D. and Spence, L. (2013). *Corporate Social Responsibility: Readings and Cases in a Global Context*. London: Routledge.

Crane, T. (2001). *The Elements of Mind*. Oxford: Oxford University Press.

Cushman, F. (2008). Crime and Punishment: Distinguishing the Roles of Causal and Intentional Analyses in Moral Judgment. *Cognition 108* (2), 353-380.

Dan-Cohen, M. (1986). *Rights, Persons, and Organizations*. Berkeley: University of California Press.

Davidson, D. (1980a). Causal Relations. In D. Davidson, *Essays on Actions and Events* (pp. 149-162). Oxford: Clarendon Press.

Davidson, D. (1980b). Mental Events. In D. Davidson, *Essays on Actions and Events* (pp. 207-227). Oxford University Press.

Davidson, D. (1980c). *Essays on Actions and Events.* Oxford: Clarendon Press.

Dennett, D. (1987). *The Intentional Stance*. Cambridge: MIT Press.

DeRose, K. (1995). Solving the Skeptical Problem. *Philosophical Review 104* (1), 1-52.

Dorsey, D. (2012). Consequentialism, Metaphysical Realism and the Argument from Cluelessness. *Philosophical Quarterly 62* (246), 48-70.

Dowe, P. (1992). Wesley Salmon's Process Theory of Causality and the Conserved Quantity Theory. *Philosophy of Science 59* (2), 195-216.

Dowe, P. (2000). *Physical Causation*. Cambridge: Cambridge University Press.

Dowe, P. (2001). A Counterfactual Theory of Prevention and 'Causation' by Omission. *Australasian Journal of Philosophy 79* (2), 216-226.

Dowe, P. (2004). Causes Are Physically Connected to Their Effects. In C. Hitchcock (Ed.), *Contemporary Debates in Philosophy of Science* (pp. 189-196). Malden: Blackwell Publishers.

Dreier, J. (1990). Internalism and Speaker Relativism. *Ethics 101* (1), 6-26.

Driver, J. (2008). Attributions of Causation and Moral Responsibility. In W. Sinnott-Armstrong (Ed.), *Moral Psychology*, Vol. 2 (pp. 426-439). Cambridge: MIT Press.

Ducasse, C. J. (1926). On the Nature and the Observability of the Causal Relation. *Journal of Philosophy 23* (3), 57-68.

Duff, A. (2007). *Answering for Crime. Responsibility and Liability in the Criminal Law*. Oxford: Hart Publishing.

Duff, A. (2009). Legal and Moral Responsibility. *Philosophy Compass 4* (6), 978-986.

Easton, D. (1953). *The Political System: An Inquiry into the State of Political Science*. New York: Knopf.

Edgarton, H. (1924). Legal Cause. *University of Pennsylvania Law Review 72*, 211-244.

Ehring, D. (1986). The Transference Theory of Causation. *Synthese 67* (2), 249-258.

Ehring, D. (1997). *Causation and Persistence*. Oxford: Oxford University Press.

Ehring, D. (2009). Causal Relata. In H. Beebee, C. Hitchcock, and P. Menzies (Eds.), *The Oxford Handbook of Causation* (pp. 387-413). Oxford: Oxford University Press.

Erskine, T. (2003). *Can Institutions Have Responsibilities? Collective Moral Agency and International Relations*. Houndmills: Palgrave Macmillan.

Esfeld, M. (2009). The Modal Nature of Structures in Ontic Structural Realism. *International Studies in the Philosophy of Science 23* (2), 179-194.

Fair, D. (1979). Causation and the Flow of Energy. *Erkenntnis 14* (3), 219-250.

Fantl, J. and McGrath, M. (2009). *Knowledge in an Uncertain World*. Oxford: Oxford University Press.

Feinberg, J. (1970). *Doing and Deserving: Essays in the Theory of Responsibility*. Princeton: Princeton University Press.

Feldman, F. (1997). *Utilitarianism, Hedonism, and Desert: Essays in Moral Philosophy.* Cambridge: Cambridge University Press.

Field, H. (2003). Causation in the Physical World. In M. Loux and D. Zimmerman (Eds.), *The Oxford Handbook of Metaphysics* (pp. 435-460). Oxford: Oxford University Press.

Fischer, J. M. and Ravizza, M. (1994). Responsibility and History. *Midwest Studies in Philosophy 19* (1), 430-451.

Fischer, J. M. and M. Ravizza (1998). *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press.

Fodor, J. A. (1997). Special Sciences: Still Autonomous After All These Years. *Philosophical Perspectives 11* (11), 149-163.

Frankfurt, H. (1969). Alternate Possibilities and Moral Responsibility. *Journal of Philosophy 66* (3), 829-839.

Franklin-Hall, L. R. (forthcoming-a). Explaining Causal Selection with Explanatory Causal Economy: Biology and Beyond. In P.-A. Braillard and C. Malaterre (Eds.), *Explanation in Biology: An Enquiry into the Diversity of Explanatory Patterns in the Life Sciences*. Dordrecht: Springer.

Franklin-Hall, L. R. (forthcoming-b). The Causal Economy Approach to Scientific Explanation. In C.K. Waters (Ed.), *Minnesota Studies in the Philosophy of Science.* Minneapolis: University of Minnesota Press.

French, P. (1979). The Corporation as a Moral Person. *American Philosophical Quarterly 16* (3), 207-215.

French, P. A. (1984). *Collective and Corporate Responsibility*. New York: Columbia University Press.

French, P. A. (1995). *Corporate Ethics*. Forthworth: Harcourt Brace.

Frey, R. G. (1974). On Causal Consequences. *Canadian Journal of Philosophy 4* (2), 365-379.

Frey, R. G. (1976). Judgments of Causal Importance in the Social Sciences. *Philosophy of the Social Sciences 6* (3), 245-248.

Frey, R. G. (1978a). Contributory Causation and the Objectivity of the Social Sciences. *Philosophy of the Social Sciences 8* (2), 175-179.

Frey, R. G. (1978b). Contributory Causation and Objectivity: A Final Instalment. *Philosophy of the Social Sciences 8* (2), 182-183.

Frey, R. G. (1978c). Causal Responsibility and Contributory Causation. *Philosophy and Phenomenological Research 39* (1), 106-119.

Funkhouser, E. (2002). Three Varieties of Causal Overdetermination. *Pacific Philosophical Quarterly 83* (4), 335-351.

Ganeri, J. Noordhof, P. and Ramachandran, M. (1996). Counterfactuals and Preemptive Causation. *Analysis 56* (4), 219-225.

Gardner, J. (2012). Wrongdoing by Results: Moore's Experiential Argument. *Legal Theory 1* (1), 1-13.

Garfinkel, A. (1981). *Forms of Explanation: Rethinking the Questions in Social Theory*. New Haven: Yale University Press.

Gasking, D. (1955). Causation and Recipes. *Mind 64* (256), 479-487.

Gilbert, M. (2009). Shared Intention and Personal Intentions. *Philosophical Studies 144* (1), 167-187.

Gilbert, M. (2013). *Joint Commitment: How We Make the Social World.* New York: Oxford University Press.

Godfrey-Smith, P. (2009). Causal Pluralism. In H. Beebee, C. Hitchcock, and P. Menzies (Eds.), *The Oxford Handbook of Causation* (pp. 325-337). Oxford: Oxford University Press.

Good, I. J. (1961). A Causal Calculus I. *British Journal for the Philosophy of Science 11* (44), 305-318.

Good, I. J. (1962). A Causal Calculus II. *British Journal for the Philosophy of Science 12* (45), 43-51.

Gorovitz, S. (1965). Causal Judgments and Causal Explanations. *Journal of Philosophy 62* (23), 695-711.

Greco, J. (2008). What's Wrong with Contextualism? *Philosophical Quarterly 58* (232), 416-436.

Grice, H. P. (1975). Logic and Conversation. In P. Cole and J. L. Morgan (Eds.), *Syntax and Semantics*, Vol. 3, Speech Acts (pp. 41-58). New York: Academic Press.

Haidt, J., Bjoklund F. (2006). Social Intuitionists Reason, in Conversation. In W. Sinnott-Armstrong (Ed.), *Moral Psychology*, Vol. 2 (pp. 426-439). Cambridge: MIT Press.

Hall, N. (2004). Two Concepts of Causation. In J. Collins, N. Hall and L. A. Paul (Eds.), *Causation and Counterfactuals* (pp. 181-276). Cambridge: MIT Press.

Hall, N. (2007). Structural Equations and Causation. *Philosophical Studies 132* (1), 109-136.

Halpern, J. Y. (2008). Defaults and Normality in Causal Structures. In G. Brewka and J. Lang (Eds.), *Principles of Knowledge Representation and Reasoning: Proceedings of the Eleventh International Conference* (pp. 198-208). Menlo Pak: AAAI Press.

Halpern, J. Y. (2013). Appropriate Causal Models and Stability of Causation. Unpublished manuscript.

Halpern, J. Y. and Pearl, J. (2005). Causes and Explanations: A Structural-Model Approach. Part I: Causes. *British Journal for the Philosophy of Science 56* (4), 843-887.

Halpern, J. Y. and Hitchcock, C. (2010). Actual Causation and the Art of Modeling. In R. Dechter, H. Geffner, and J. Y. Halpern (Eds.), *Heuristics, Probability and Causality: A Tribute to Judea Pearl* (pp. 383-406). London: College Publications.

Halpern, J. Y. and Hitchcock, C. (2013). Compact Representations of Extended Causal Models. *Cognitive Science 37* (6), 986-1010.

Halpern, J. Y. and Hitchcock, C. (forthcoming). Graded Causation and Defaults. *British Journal for the Philosophy of Science*.

Hankinson, R. J. (1998). *Cause and Explanation in Ancient Greek Thought.* Oxford: Clarendon Press.

Hanson, N. R. (1958). *Patterns of Discovery.* Cambridge: Cambridge University Press.

Hart, H. L. A. (1961). *The Concept of Law*. Oxford: Clarendon Press.

Hart, H. L. A. and Honoré, T. (1985). *Causation in the Law*. Oxford: Clarendon Press.

Hawthorne, J. (2001). Freedom in Context. *Philosophical Studies 104* (1), 63-79.

Hawthorne, J. (2004). *Knowledge and Lotteries.* Oxford: Clarendon Press.

Heinaman, R. (1986). Incompatibilism Without the Principle of Alternative Possibilities. *Australasian Journal of Philosophy 64* (3), 266-276.

Hesslow, G. (1983). Explaining Differences and Weighting Causes. *Theoria 49* (2), 87-111.

Hesslow, G. (1988). The Problem of Causal Selection. In D. J. Hilton (Ed.), *Contemporary Science and Natural Explanation: Commonsense Conceptions of Causality* (pp. 11-31). New York: New York University Press.

Hitchcock, C. (1995). Salmon on Explanatory Relevance. *Philosophy of Science 62* (2), 304-320.

Hitchcock, C. (1996). The Role of Contrast in Causal and Explanatory Claims. *Synthese 107* (3), 395-419.

Hitchcock, C. (1999). Contrastive Explanation and the Demons of Determinism. *British Journal for the Philosophy of Science 50* (4), 585-612.

Hitchcock, C. (2001). The Intransitivity of Causation Revealed in Equations and Graphs. *Journal of Philosophy 98* (6), 273-299.

Hitchcock (2007a). Three Concepts of Causation. *Philosophy Compass 2* (3), 508-516.

Hitchcock (2007b). Prevention, Preemption, and the Principle of Sufficient Reason. *Philosophical Review 116* (4), 495-532.

Hitchcock, C. and Knobe, J. (2009). Cause and Norm. *The Journal of Philosophy 106* (11), 587-612.

Hobbes, T. (1994). *Leviathan* (edited by E. Curley). Indianapolis: Hackett.

Honoré, T. (1988). Responsibility and Luck. *Law Quarterly Review 104*, 530-553.

Horgan, T. E. (1989). Mental Quausation. *Philosophical Perspectives 3*, 47-74.

Horgan, T. and Timmons, M. (2002). Conceptual Relativity and Metaphysical Realism. *Noûs* 36 (s1), 74-96.

Horwich, P. (1987). *Asymmetries in Time*. Cambridge: MIT Press.

Hume, D. (1739/2001). *A Treatise of Human Nature*. Oxford: Oxford University Press.

Hunt, I. (2005). Omissions and Preventions as Cases of Genuine Causation. *Philosophical Papers 34* (2), 209-233.

Jackson, F. (1991). Decision-Theoretic Consequentialism and the Nearest and Dearest Objection. *Ethics 101* (3), 461-482.

James, W. (1980/1981). *The Principles of Psychology*. Cambridge: Harvard University Press.

Jenkins, C. S. and Nolan, D. (2010). Maximising, Satisficing and Context. *Noûs 44* (3), 451-468.

Kagan, S. (1998). *Normative Ethics*. Boulder: Westview Press.

Keeley, M. (1981). Organizations as Non-Persons. *Journal of Value Inquiry 15* (2), 149-155.

Kelman, M. (1987). The Necessary Myth of Objective Causation Judgments in Liberal Political Theory. *Chicago-Kent Law Review 63*, 603-604.

Kim, J. (1973). Causation, Nomic Subsumption, and the Concept of Event. *Journal of Philosophy 70* (8), pp. 217-236

Kim, J. (1974). Noncausal Connections. *Noûs 8* (1), 41-52.

Kim, J. (1976). Events as Property Exemplifications. In M. Brand and D. Walton (Eds.), *Action Theory* (pp. 310-326). Dordrecht: D. Reidel.

Kim, J. (1981). Causes as Explanations: A Critique. *Theory and Decision 13* (4), 293-309.

Kim, J. (1989). The Myth of Nonreductive Materialism. *Proceedings of the American Philosophical Association 63* (3), 31-47.

Kim, J. (1993a). Mechanism, Purpose, and Explanatory Exclusion. In J. Kim (1993c) *Supervenience and Mind* (pp. 237-264). Cambridge: Cambridge University Press.

Kim, J. (1993b). The Nonreductivist's Troubles with Mental Causation. In J. Kim (1993c) *Supervenience and Mind* (pp. 336-357). Cambridge: Cambridge University Press.

Kim, J. (1993c) *Supervenience and Mind*. Cambridge: Cambridge University Press.

Kim, J. (1993d). Explanatory Realism, Causal Realism, and Explanatory Exclusion. In D-H. Ruben (Ed.), *Explanation* (pp. 229-245). Oxford: Oxford University Press.

Kistler, M. (1998). Reducing Causality to Transmission. *Erkenntnis 48* (1), 1-25.

Kornhauser, L. and Sager, L. (1993). The One and the Many: Adjudication in Collegial Courts. *California Law Review 81*, 1-59.

Kurki, M. (2008). *Causation in International Relations: Reclaiming Causal Analysis*. Cambridge: Cambridge University Press.

Kutz, C. (2002). *Complicity. Ethics and Law for a Collective Age*. Cambridge: Cambridge University Press.

Kvart, I. (1997). Cause and Some Positive Causal Impact. In J. Tomberlin (Ed.), *Philosophical Perspectives 11: Mind, Causation, and World* (pp. 401-432). Oxford: Basil Blackwell.

Kvart, I. (2004). Probabilistic Cause, Edge Conditions, Late Preemption and Discrete Cases. In P. Dowe and P. Noordhof (Eds.), *Cause and Chance: Causation in an Indeterministic World* (pp. 163-188). London: Routledge.

Ladd, J. (1970). Morality and the Ideal of Rationality in Formal Organizations. *The Monist 54* (4), 488-516.

Landes J. and Posner, R (1983). Causation in Tort Law: An Economic Approach. *Journal of Legal Studies 12*, 109-134.

Lenman, J. (2000). Consequentialism and Cluelessness. *Philosophy and Public Affairs 29* (4), 342-370.

LePore, E. and Loewer, B. M. (1987). Mind Matters. *Journal of Philosophy 84* (11), 630-642.

Lewis, D. (1973). Causation. *The Journal of Philosophy 70* (17), 556-567.

Lewis, D. (1986a). Causal Explanation. In D. Lewis (Ed.), *Philosophical Papers Vol. II* (pp. 214-240). Oxford: Oxford University Press.

Lewis, D. (1986b). Postscripts to 'Causation'. In D. Lewis (Ed.), *Philosophical Papers* Vol. II (pp. 172-213). Oxford: Oxford University Press.

Lewis, D. (1996). Elusive Knowledge. *Australasian Journal of Philosophy 74* (4), 549-567.

Lewis, D. (2000). Causation as Influence. *Journal of Philosophy 97* (4), 182-197.

Lewis, D. (2004). Void and Object. In J. Collins, N. Hall and L. A. Paul (Eds.), *Causation and Counterfactuals* (pp. 277-290). Cambridge: MIT Press.

Lipton, P. (1990). Contrastive Explanation. In D. Knowles (Ed.), *Explanation and its Limits*, Royal Institute of Philosophy Supplement, Vol. 27 (pp. 247-266). Cambridge: Cambridge University Press.

Lipton, P. (1991). *Inference to the Best Explanation*. London: Routledge.

List, C. and Pettit, P. (2006). Group Agency and Supervenience. *Southern Journal of Philosophy 44* (1), 85-105.

List, C. and Pettit, P. (2011). *Group Agency: The Possibility, Design, and Status of Corporate Agents.* Oxford: Oxford University Press.

List, C. and Spiekermann, K. (2013). Methodological Individualism and Holism in Political Science: A Reconciliation. *American Political Science Review 107* (4), 629-643.

Livengood, J., Sytsma, J. and Rose, D. (forthcoming). Following the FAD: Folk Attributions and Theories of Actual Causation.

Locke, J. (1960). *Two Treatises of Government*. Cambridge: Cambridge University Press.

Ludwig, K. (2007). Normative Autonomy for Collective Agents. *Journal of Social Philosophy 38* (3), 410-427.

Mackie, J. L. (1955). Responsibility and Language. *Australasian Journal of Philosophy 33* (3), 143-159.

Mackie, J. L. (1965). Causes and Conditions. *American Philosophical Quarterly 2* (4), 245-264.

Mackie, J. L. (1980). *The Cement of the Universe: A Study of Causation*. Oxford: Clarendon Press.

Malone, W. (1956). Ruminations on Cause-In-Fact. *Stanford Law Review 9*, 60-99.

Martin, M. (1974). Causal Importance and Objectivity. *Philosophy of the Social Sciences 4* (2), 157-168.

Martin, M. (1978a). Judgements of Contributory Causes and Objectivity. *Philosophy of the Social Sciences 8* (2), 173-175.

Martin, M. (1978b). Contributory Causes Again. *Philosophy of the Social Sciences 8* (2), 180-181.

Maslen, C. (2004). Causes, Contrasts, and the Nontransitivity of Causation. In J. Collins, E. Hall, and L. A. Paul (Eds.), *Causation and Counterfactuals* (pp. 341-358). Cambridge: MIT Press.

May, S. K., Cheney, G. and Roper, J. (2007). *The Debate over Social Corporate Responsibility*. Oxford: Oxford University Press.

McDermott, M. (1995). Redundant Causation. *British Journal for the Philosophy of Science 46* (4), 523-544.

McDermott, M. (2002). Causation: Influence versus Sufficiency. *Journal of Philosophy 99* (2), 84-101.

McGill, A. L. (1991). The Influence of the Causal Background on the Selection of Causal Explanations. *British Journal of Social Psychology 30*, 79-87.

McGrath, S. (2002). The Causal Relata. Unpublished manuscript.

McGrath, S. (2005). Causation by Omission: A Dilemma. *Philosophical Studies 123* (1-2), 125-148.

McLaughlin, J. A. (1925). Proximate Cause. *Harvard Law Review 39* (2), 149-199.

Mele, A. (2009). Causation, Action, and Free Will. In H. Beebee, C. Hitchcock, and P. Menzies (Eds.), *The Oxford Handbook of Causation* (pp. 554-574). Oxford: Oxford University Press.

Mellor, D. H. (1995). *The Facts of Causation*. London: Routledge.

Mellor, D. H. (2004). For Facts as Causes and Effects. In J. Collins, E. Hall, and L. A. Paul (Eds.), *Causation and Counterfactuals* (pp. 309-324). Cambridge: MIT Press.

Menzies, P. (1989). Probabilistic Causation and Causal Processes: A Critique of Lewis. *Philosophy of Science 56* (4), 642-663.

Menzies, P. (1996). Probabilistic Causation and the Pre-Emption Problem. *Mind 105* (417), 85-117.

Menzies, P. (2004). Difference-Making in Context. In J. Collins, E. Hall, and L. A. Paul (Eds.), *Causation and Counterfactuals* (pp. 139-180). Cambridge: MIT Press.

Menzies, P. (2007). Causation in Context. In H. Price and R. Corry (Eds.), *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited* (pp 191-223). Oxford: Clarendon Press.

Menzies, P. (2009). Platitudes and Counterexamples. In H. Beebee, C. Hitchcock, and P. Menzies (Eds.), *The Oxford Handbook of Causation* (pp. 341-367). Oxford: Oxford University Press.

Menzies, P. and Price, H. (1993). Causation as a Secondary Quality. *The British Journal for the Philosophy of Science 44* (2), 187-203.

Menzies, P. and List, C. (2010). The Causal Autonomy of the Special Sciences. In C. Mcdonald and G. Mcdonald (Eds.), *Emergence in Mind* (pp. 108-128). Oxford: Oxford University Press.

Merricks, T. (2001) *Objects and Persons*. Oxford: Oxford University Press.

Mill, J. S. (1843/1947). *A System of Logic*. London: Longmans, Green and Co.

Miller, S. (2002). Against Collective Agency. In G. Meggle (Ed.), *Social Facts and Collective Intentionality* (pp. 273-298). Frankfurt: Dr. Hansel-Hohenhausen Ag.

Montminy, M. (2007). Moral Contextualism and the Norms for Moral Conduct. *American Philosophical Quarterly 44* (1), 1-13.

Moore, G. E. (1903). *Principia Ethica*. Cambridge: Cambridge University Press.

Moore, G. E. (1942). A Reply to My Critics. In P. A. Schilpp (Ed.), *The Philosophy of G. E. Moore* (pp. 535-677). Evanston: Northwestern University Press.

Moore, M. S. (2009). *Causation and Responsibility*. Oxford: Oxford University Press.

Moore, M. S. (2012). Moore's Truths about Causation and Responsibility: A Reply to Alexander and Ferzan. *Criminal Law and Philosophy 6*, 445-462.

Morse, S. J. (2000). The Moral Metaphysics of Causation and Results. *California Law Review 88* (3), 879-894.

Morse, S. J. (2004). Reason, Results, and Criminal Responsibility. *Illinois Law Review 2*, 363-444.

Moya, C. J. (2006). *Moral Responsibility: the Ways of Scepticism*. London: Routledge.

Mumford, S. and Anjum, R. L. (2011). *Getting Causes from Powers*. Oxford: Oxford University Press.

Mumford, S. (forthcoming). Causes for Laws. *Jurisprudence 3*.

Nagel, T. (1979). Moral Luck. In T. Nagel *Mortal Questions* (pp. 24-38). Cambridge: Cambridge University Press.

Nisbett, R. E. and Cohen, D. (1996). *Culture of Honour: The Psychology of Violence in the South*. Boulder: Westview Press.

Norcross**, A.** (1990). Consequentialism and the Unforeseeable Future. *Analysis 50* (4), 253-256.

Norcross, A. (2005a). Contextualism for Consequentialists. *Acta Analytica 20* (2), 80-90.

Norcross, A. (2005b). Harming in Context. *Philosophical Studies 123* (1-2), 149-173.

Northcott, R. (2008). Causation and Contrast Classes. *Philosophical Studies 139* (1), 111-123.

Northcott, R. (2013a). Verisimilitude: A Causal Approach. *Synthese 190* (9), 1471-1488.

Northcott, R. (2013b). Degree of Explanation. *Synthese 190* (15), 3087-3105.

Paul, L. A. (2000). Aspect Causation. *Journal of Philosophy 97* (4), 235-256.

Paul, L. A. (2009). Counterfactual Theories. In H. Beebee, C. Hitchcock, and P. Menzies (Eds.), *The Oxford Handbook of Causation* (pp. 158-184). Oxford: Oxford University Press.

Paul, L. A. and Hall, N. (2013). *Causation: A User's Guide*. Oxford: Oxford University Press.

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.

Persson, J. (2002). Cause, Effect, and Fake Causation. *Synthese 131* (1), 129-143.

Pettit, P. (1993). *The Common Mind: An Essay on Psychology, Society, and Politics*. New York: Oxford University Press.

Pettit, P. (2001). *A Theory of Freedom: From the Psychology to the Politics of Agency*. New York: Oxford University Press.

Pettit, P. (2003). Groups with Minds of Their Own. In F. Schmidtt (Ed.), *Socializing Metaphysics: The Nature of Social Reality* (pp. 167-193). Lanham: Rowman and Littlefield.

Pettit, P. (2007). Responsibility Incorporated. *Ethics 117* (2), 171-201.

Pettit, P. (2009). The Reality of Group Agents. In C. Mantzavinos (Ed.), *Philosophy of the Social Sciences: Philosophical Theory and Scientific Practice* (pp. 67-91). Cambridge: Cambridge University Press.

Pettit, P. and Schweikard, D. (2006). Joint Actions and Group Agents. *Philosophy of the Social Sciences, 36* (1), 18-39.

Pettit, P. (forthcoming). Group Agents are Not Expressive, Pragmatic, or Theoretical Fictions.

Plutarch (1984). *Life of Pericles*. Translated by H. Holden. Bolchazy-Carducci Publishers.

Price, H. and Corry, R. (2007). *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*. Oxford: Clarendon Press.

Psillos, S. (2002). *Causation and Explanation*. Chesham: Acumen.

Putnam, H. (1984). Is the Causal Structure of the Physical Itself Something Physical? *Midwest Studies in Philosophy 9* (1), 3-16.

Quine, W. V. O. (1960). *Word and Object*. Cambridge: MIT Press.

Quine, W. V. O. (1966). *The Ways of Paradox*. New York: Random House.

Quinton, A. (1976). Social Objects. *Proceedings of the Aristotelian Society 76*, 1-27.

Raatikainen, P. (2006) Mental Causation, Interventions, and Contrasts. Unpublished manuscript.

Railton, P. (1984). Alienation, Consequentialism, and the Demands of Morality. *Philosophy and Public Affairs 13* (2), 134-171.

Raz, J. (1982). Promises in Morality and Law. *Harvard Law Review 95*, 916-38.

Reichenbach, H. (1949). *The Theory of Probability. An Inquiry into the Logical and Mathematical Foundations of the Calculus of Probability*. Berkeley and Los Angeles: University of California Press.

Reiss, J. (2009). Causation in the Social Sciences: Evidence, Inference, and Purpose. *Philosophy of the Social Sciences 39* (1), 20-40.

Rieber, S. (2006). Free Will and Contextualism. *Philosophical Studies 129* (2), 223-252.

Risser, D. T. (2006). Collective Moral Responsibility. *Internet Encyclopedia of Philosophy*, http://www.iep.utm.edu/collecti/

Robinson, P. H. and Darley, J. M. (1997). The Utility of Desert. *Northwestern University Law Review 91*, 453-99.

Rosen, G. (2004). Skepticism about Moral Responsibility. *Philosophical Perspectives 18* (1), 295-313.

Rousseau, J.-J. (1997). *The Social Contract and Later Political Writings* (edited by V. Gourevitch). Cambridge: Cambridge University Press.

Rousseau, J.-J. (1967). Rousseau to Voltaire, 18 August 1756. In J.A. Leigh (Ed.), *Correspondence Complète de Jean Jacques Rousseau*, Vol. 4 (pp. 37-50). Translated by R. Spang. Geneva: Institute et Musée Voltaire les Délices.

Rovane, C. (1997). *The Bounds of Agency: An Essay in Revisionary Metaphysics*. Princeton: Princeton University Press.

Ruddick, W. (1968). Causal Connection. *Synthese 18* (1), 46-67.

Rueger, A. (2006). Connection and Influence: A Process Theory of Causation. *Journal for General Philosophy of Science 37* (1), 77-97.

Runciman, D. (1997). *Pluralism and the Personality of the State*. Cambridge: Cambridge University Press.

Russell, B. (1913). On the Notion of Cause. *Proceedings of the Aristotelian Society 13*, 1-26.

Russell, B. (1948). *Human Knowledge: Its Scope and Limits*. New York: Simon and Schuster.

Rysiew, P. (2001). The Context-Sensitivity of Knowledge Attributions. *Noûs 35* (4), 477-514.

Salmon, W. (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.

Salmon, W. (1994). Causality without Counterfactuals. *Philosophy of Science 61* (2), 297-312.

Salmon, W. (1997). Causality and Explanation: A Reply to Two Critiques. *Philosophy of Science 64* (3), 461-477.

Salmon, W. (1998). *Causality and Explanation*. New York: Oxford University Press.

Saramago, J. (1994). *The Gospel According to Jesus Christ*. New York: Harcourt Brace.

Sartorio, C. (2004). How to Be Responsible for Something Without Causing It. *Philosophical Perspectives 18* (1), 315-336.

Sartorio, C. (2005a). A New Asymmetry between Actions and Omissions. *Noûs 39* (3), 460-482.

Sartorio, C. (2005b). Causes as Difference-Makers. *Philosophical Studies 123* (1-2), 71-96.

Sartorio, C. (2006a). On Causing Something to Happen in a Certain Way without Causing It to Happen. *Philosophical Studies 129* (1), 119-136.

Sartorio, C. (2006b). Disjunctive Causes. *Journal of Philosophy 103* (10), 521-538.

Sartorio, C. (2007). Causation and Responsibility. *Philosophy Compass 2* (5), 749-765.

Sartorio, C. (2009a). Omissions and Causalism. *Noûs 43* (3), 513-530.

Sartorio, C. (2009b). Causation in Ethics. In H. Beebee, C. Hitchcock, and P. Menzies (Eds.), *The Oxford Handbook of Causation* (pp. 575-591). Oxford: Oxford University Press.

Sartorio, C. (2010). The Prince of Wales Problem for Counterfactual Theories of Causation. In A. Hazzlett (Ed.), *New Waves in Metaphysics* (pp. 259-276). Houndmills: Palgrave Macmillan.

Sartorio, C. (2011). Moore on Doing versus Allowing Harm. *Rutgers Law Journal 42* (2), 435-449.

Sartorio, C. (2012). Resultant Luck. *Philosophy and Phenomenological Research 84* (1), 63-86.

Sartorio, C. (forthcoming). The Problem of Determinism and Free Will Is Not the Problem of Determinism and Free Will. In A. Mele (Ed.), *Surrounding Free Will*. New York: Oxford University Press.

Schaffer, J. (2000). Causation by Disconnection. *Philosophy of Science 67* (2), 285-300.

Schaffer, J. (2001). Causes as Probability Raisers of Processes. *Journal of Philosophy 98* (2), 75-92.

Schaffer, J. (2003). Overdetermining Causes. *Philosophical Studies 114* (1-2), 23-45.

Schaffer, J. (2004). Causes Need Not Be Physically Connected to Their Effects: The Case for Negative Causation. In C. Hitchcock (Ed.), *Contemporary Debates in Philosophy of Science* (pp. 197-216). Malden: Basil Blackwell.

Schaffer, J. (2005). Contrastive Causation. *Philosophical Review 114* (3), 327-358.

Schaffer, J. (2007). The Metaphysics of Causation. *Stanford Encyclopaedia of Philosophy*, http://plato.stanford.edu/entries/causation-metaphysics/

Schaffer, J. (2008). Knowledge in the Image of Assertion. *Philosophical Issues 18* (1), 1-19.

Schaffer, J. (2010). Contrastive Causation in the Law. *Legal Theory 16* (4), 259-297.

Schaffer, J. (2012). Causal Contextualisms. In M. Blaauw (Ed.), *Contrastivism in Philosophy: New Perspectives* (pp. 35-63). London: Routledge.

Schroeder, M. (2009). Hybrid Expressivism: Virtues and Vices. *Ethics 119* (2), 257-309.

Schweder, R. (1999). Causal Explanation and Explanatory Selection. *Synthese 120* (1), 115-124.

Scriven, M. (1975). Causation as Explanation. *Noûs 9* (1), 3-16.

Searle, J. (1995). *The Construction of Social Reality*. New York: Free Press.

Shorter, J. M. (1965). Causality: A Method of Analysis. In R. J. Butler (Ed.), *Analytical Philosophy* (pp. 145-157). Oxford: Blackwell.

Sider, T. (2003). What's So Bad About Overdetermination? *Philosophy and Phenomenological Research 67* (3), 719-726.

Sinnott-Armstrong, W. (2006a). *Moral Skepticisms*. Oxford: Oxford University Press.

Sinnott-Armstrong, W. (2006b). Framing Moral Intuitions. In W. Sinnott-Armstrong (Ed.), *Moral Psychology*, Vol. 2 (pp. 47-76). Cambridge: MIT Press.

Sinnott-Armstrong, W. (2008). Summary of *Moral Skepticisms*. *Philosophical Books 49* (3), 193-196.

Skyrms, B. (1984). EPR: Lessons for Metaphysics. *Midwest Studies in Philosophy 9* (1), 245-255.

Smart, J. J. C. and Williams, B. (1973). *Utilitarianism: For and Against*. London: Cambridge University Press.

Sophocles (1984). *Oedipus at Colonus*. Translated by R. Fitzgerald. New York: Harcourt, Brace and World.

Sosa, D. (1993). Consequences of Consequentialism. *Mind 102* (405), 101-122.

Spencer, H. (1860). The Social Organism. *Westminster Review 73*, 51-68.

Spirtes, P., Glymour, C. and Scheines, R. (1993). *Causation, Prediction, and Search*. New York: Springer-Verlaag.

Spohn, W. (1993). Causal Laws Are Objectifications of Inductive Schemes. In J. Dubucs (Ed.), *Philosophy of Probability* (pp. 223-252). Dordrecht: Kluwer.

Stanley, J. (2005). *Knowledge and Practical Interests.* Oxford University Press.

Stapleton, J. (2008). *Choosing what we mean* by 'Causation' in the Law. *Missouri Law Review 73* (2), 433-480.

Stapleton, J. (2009). Causation in the Law. In H. Beebee, C. Hitchcock, and P. Menzies (Eds.), *The Oxford Handbook of Causation* (pp. 744-769). Oxford: Oxford University Press.

Steward, H. (1997). On the Notion of Cause 'Philosophically Speaking'. *Proceedings of the Aristotelian Society 97* (2), 125-140.

Steward, H. (2008). Fresh Starts. *Proceedings of the Aristotelian Society 108* (1), 197-217.

Strawson, G. (1989). *The Secret Connexion*. Oxford: Clarendon Press.

Strawson, G. (1994). The Impossibility of Moral Responsibility. *Philosophical Studies 75* (1-2), 5-24.

Strevens, M. (2008). *Depth: An Account of Scientific Explanation*. Cambridge: Harvard University Press.

Strevens, M. (2013). Causality Reunified. *Erkenntnis 78* (2), 299-320.

Suárez, A. (2014). The Contextual Character of Causal Evidence. *Topoi 33* (2), 397-406.

Suppes, P. (1970). *A Probabilistic Theory of Causality*. Amsterdam: North Holland Publishing.

Suppes, P. (1984). Conflicting Intuitions about Causality. *Midwest Studies in Philosophy 9* (1), 151-168.

Swain, M. (1978). A Counterfactual Analysis of Event Causation. *Philosophical Studies 34* (1), 1-19.

Swanson, E. (2010). Lessons from the Context Sensitivity of Causal Talk. *The Journal of Philosophy 107* (5), 221-242.

Swanson, E. (forthcoming). The Language of Causation. In D. G. Fara and G. Russell (Eds.), *The Routledge Companion to the Philosophy of Language*. London: Routledge.

Thomson, J. J. (2003). Causation: Omissions. *Philosophy and Phenomenological Research, 66* (1), 81-103.

Tollefsen, D. P. (2002a). Collective Intentionality and the Social Sciences. *Philosophy of the Social Sciences 32* (1), 25-50.

Tollefsen, D. (2002b). Organizations as True Believers. *Journal of Social Philosophy 33* (3), 395-410.

Tollefsen, D. P. (2003). Participant Reactive Attitudes and Collective Responsibility. *Philosophical Explorations 6* (3), 218-234.

Tooley, M. (1987). *Causation: A Realist Approach*. Oxford: Clarendon Press.

Tooley, M. (2004). Probability and Causation. In P. Dowe and P. Noordhof (Eds.), *Cause and Chance: Causation in an Indeterministic World* (pp. 77-119). London: Routledge.

Tuomela, R. (1995). *The Importance of Us*. Stanford: Stanford University Press.

Tuomela, R. (2005). We-intentions Revisited. *Philosophical Studies 125* (3), 327-369.

Tuomela, R. (2007). *The Philosophy of Sociality: The Shared Point of View*. New York: Oxford University Press.

Unger, P. (1995). Contextual Analysis in Ethics. *Philosophy and Phenomenological Research 55* (1), 1-26.

van Fraassen, Bas (1980). *The Scientific Image*. Oxford: Oxford University Press.

Varzi, A. (2007). Omissions and Causal Explanations. In F. Castellani and J. Quitterer (Eds.), *Agency and Causation in the Human Sciences* (pp. 155-167). Paderborn: Mentis Verlag.

Vogel, D. (2006). *The Market for Virtue. The Potential and Limits of Corporate Social Responsibility*. Washington: Brookings Institution.

Voltaire (1911). *Poème sur le désastre de Lisbonne* (1755). In J. McCabe (Ed.), *Selected Works of Voltaire* (pp. 10-14). London: Watts and Co.

von Wright, G. H. (1971). *Explanation and Understanding*. Ithaca: Cornell University Press.

von Wright, G. H. (1975). On the Logic and Epistemology of the Causal Relation. In E. Sosa (Ed.), *Causation and Conditionals* (pp. 95-113). Oxford: Oxford University Press.

Waters, C. K. (2007). Causes That Make a Difference. *Journal of Philosophy 104* (11), 551-579.

Watson, G. (1996). Two Faces of Responsibility. *Philosophical Topics 24* (2), 227-248.

Weatherson, B. (2008). Attitudes and Relativism. *Philosophical Perspectives 22* (1), 527-544.

Wedgwood, R. (2006a). The Meaning of 'Ought'. In R. Shafer-Landau (Ed.), *Oxford Studies in Metaethics*, Vol. 1 (pp. 127-160). Oxford: Oxford University Press.

Wedgwood, R. (2006b). The Context-sensitivity of 'Ought', and the Normativity of Rationality. Unpublished manuscript.

Williams, B. (1993). *Shame and Necessity*. Berkeley: University of California Press.

Williamson, J. (2006). Causal Pluralism versus Epistemic Causality. *Philosophica 77* (1), 69-96.

Wittgenstein, L. (1922). *Tractatus Logico-Philosophicus*. London: Routledge and Kegan Paul Ltd.

Woodward, J. (1984). A Theory of Singular Causal Explanation. *Erkenntnis 21* (3), 231-262.

Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.

Wright, R. W. (1985). Causation in Tort Law. *California Law Review 73*, 1735-1828.

Yablo, S. (1992). Mental Causation. *Philosophical Review 101* (2), 245-280.

Yablo, S. (2002). De Facto Dependence. *Journal of Philosophy 99* (3), 130-148.

Ylikoski, P. (2007). The Idea of Contrastive Explanandum. In J. Persson and P. Ylikoski (Eds), *Rethinking Explanation* (pp.27-42). Dordrecht: Springer.